BAYESIAN UNCERTAINTY QUANTIFICATION FOR LARGE SCALE SPATIAL INVERSE PROBLEMS

A Dissertation

by

ANIRBAN MONDAL

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2011

Major Subject: Statistics

BAYESIAN UNCERTAINTY QUANTIFICATION FOR LARGE SCALE SPATIAL INVERSE PROBLEMS

A Dissertation

by

ANIRBAN MONDAL

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Co-Chairs of Committee,	Bani K. Mallick
	Yalchin Efendiev
Committee Members,	Faming Liang
	Akhil Datta-Gupta
	Samiran Sinha
Head of Department,	Simon J. Sheather

August 2011

Major Subject: Statistics

ABSTRACT

Bayesian Uncertainty Quantification for Large Scale Spatial Inverse Problems.

(August 2011)

Anirban Mondal, B.S., University of Calcutta;

M.Stat., Indian Statistical Institute;

M.S., Michigan State University

Co–Chairs of Advisory Committee: Dr. Bani K. Mallick
Dr. Yalchin Efendiev

We considered a Bayesian approach to nonlinear inverse problems in which the unknown quantity is a high dimension spatial field. The Bayesian approach contains a natural mechanism for regularization in the form of prior information, can incorporate information from heterogeneous sources and provides a quantitative assessment of uncertainty in the inverse solution. The Bayesian setting casts the inverse solution as a posterior probability distribution over the model parameters. Karhunen-Loéve expansion and Discrete Cosine transform were used for dimension reduction of the random spatial field. Furthermore, we used a hierarchical Bayes model to inject multiscale data in the modeling framework. In this Bayesian framework, we have shown that this inverse problem is well-posed by proving that the posterior measure is Lipschitz continuous with respect to the data in total variation norm. The need for multiple evaluations of the forward model on a high dimension spatial field (e.g. in the context of MCMC) together with the high dimensionality of the posterior, results in many computation challenges. We developed two-stage reversible jump MCMC method which has the ability to screen the bad proposals in the first inexpensive stage. Channelized spatial fields were represented by facies boundaries and variogram-based spatial fields within each facies. Using level-set based approach, the

shape of the channel boundaries was updated with dynamic data using a Bayesian hierarchical model where the number of points representing the channel boundaries is assumed to be unknown. Statistical emulators on a large scale spatial field were introduced to avoid the expensive likelihood calculation, which contains the forward simulator, at each iteration of the MCMC step. To build the emulator, the original spatial field was represented by a low dimensional parameterization using Discrete Cosine Transform (DCT), then the Bayesian approach to multivariate adaptive regression spline (BMARS) was used to emulate the simulator. Various numerical results were presented by analyzing simulated as well as real data.

To my parents Monika and Ajit Kumar Mondal

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my co-advisors Prof. Bani K. Mallick and Prof. Yalchin Efendiev. Without their guidance and persistent help this dissertation would not have been possible. I would like to thank my PhD committee members, Prof. Akhil Datta-Gupta, Prof. Faming Liang and Dr. Samiran Sinha for their help and support. A special thanks goes to Prof. Michael Longnecker for bailing me out of difficulties with his invaluable advice and constant support. A big thanks goes to my friends and peers for the enjoyable time we spent together at Texas A&M and Michigan State University. Last, but not least, I express my profound appreciation to Paromita for her encouragement, understanding, love and support during the hard times of this study. The research work discussed in this dissertation is supported in parts by Research Grant NSF CMG 0724704 and by Award Number KUS-C1-016-04 made by King Abdullah University of Science and Technology (KAUST).

TABLE OF CONTENTS

Page

CHAPTER Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

FIGURE Page

FIGURE Page

CHAPTER I

INTRODUCTION

Mathematical models are studied using computer simulation in almost all areas of applied and computational mathematics. The indirect estimation of model parameters or inputs from observations constitutes an inverse problem. Such problems arise frequently in science and engineering, with applications in weather forecasting, climate prediction, chemical kinetics and oil reservoir forecasting. Quantifying the uncertainty in inputs or parameters is then essential for predictive modeling and simulation based decision making.

A physical system is often described by a forward model, which predicts some measurable features of the system given a set of model input parameters. The corresponding inverse problem consists of inferring these input parameters from a set of observations. The simplicity of this definition belies many fundamental challenges. For example, in large scale spatial inverse problems, the observed data may be very limited compared to the dimension of the unknown input spatial field. Moreover, the available observed data may be corrupted by noise and the action of the forward model may include filtering or smoothing effect. These features typically develop ill-posed inverse problems.

Classical statistical approaches have used various regularization methods to impose well-posedness of the inverse problems. The resulting deterministic problems are solved by optimization and other means; see for example Vogel (2002). Here we focus on the Bayesian approach to nonlinear inverse problems where the input is a high dimension spatial field. As described in Marzouk and Najm (2009), the

---

This dissertation follows the style of the *Journal of the Royal Statistical Society*.

Bayesian approach contains a natural mechanism for regularization in the form of prior information, can incorporate information from heterogeneous sources and provide a quantitative assessment of uncertainty in the inverse solution (e.g. Kaipio and Somersalo (2004)). Indeed, the Bayesian setting casts the inverse solution as a posterior probability distribution over the model parameters.

In this dissertation, we consider the inverse problems whose solutions are unknown functions, (say high dimensional spatial fields) (e.g. Ramsey and Silverman (2005) and Tarantola (2005)). Estimating spatial fields istead of parameters from noisy output data increases the ill-posedness of the inverse problem, as we have to estimate an infinite-dimensional spatial process from a finite amount of noisy data. So, we use various dimensionality reduction techniques in the Bayesian formulation of inverse problems, and allow the dependence of the dimensionality on both prior and the data. Furthermore, to obtain physically meaningful results, we incorporate additional information on the unknown field through spatially smoothing priors as well as additional multiscale data.

First let's discuss some applications of the inverse probelms, then we shall move to the general discussion on how to solve those inverse probelm.

I.1. Different Examples of the Inverse Problem

I.1.1. Reservoir Characterization

Subsurfaces are complex geological formations encompassing a wide range of physical and chemical heterogeneities. These heterogeneities span over multiple length scales and are impossible to describe in a deterministic fashion. The goal of reservoir characterization is to provide a stochastic model that can estimate reservoir attributes such as permeability, porosity and fluid saturation together with thier uncertianties (see Kim *et al.* (2005)). These attributes are then used as inputs model parameters by

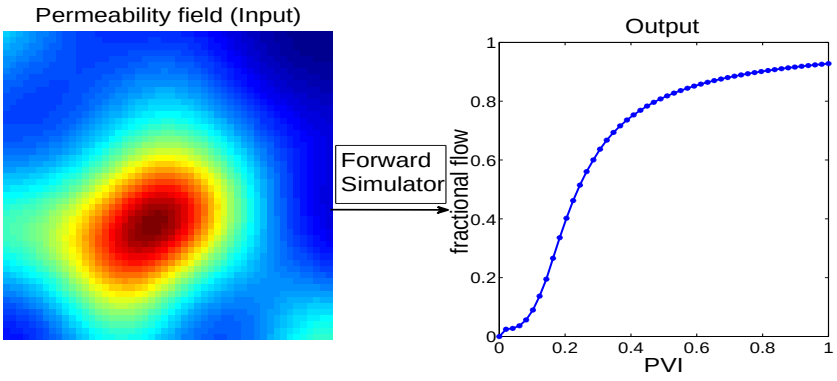

Fig. 1. The forward simulator.

various forward simulators to forecast future reservoir performance and oil recovery potential. In reservoir characterizations, the oil-water flow is typically goverened by Darcy's law where the single most influential input is the permeability spatial field, k in our notation. Permeability is an important concept in porous media flow (such as oil-water flow in reservoirs) as flow in the subsurface is controlled by the connectivity of the extreme permeabilities (high and low) which are generally associated with geological patterns that create preferential flow paths/barriers. Thus the goal of our stochastic model is to estimate the permeability field together with the uncertainties of the models. As permeability takes positive values, hence we transform $Y = log(k)$ for our modeling convenience. The main available response is the fractional flow or the water-cut data which is the fraction of water produced in relation to the total production rate in a two phase oil-water flow reservoir and denoted by d. The forward operator G which maps the water-cut data with the permeability field through a logit transformation is given by

$$d = logit[G(Y)] + \epsilon. \tag{1.1}$$

When the model input, i.e log permeability field Y is known, water-cut can be

obtained by running the forward operator G (see Figure 1), which contains several partial differential equations which has been described in Subsection II.5. We obtain permeability data in different scales. The fine-scale data represents point measurements such as well logs and cores where as the coarse-scale data can be obtained from seismic data. Our intention is to solve this inverse problem to infer about the fine-scale permeability field using the data from the output (fractional flow) and the coarse scale data. In other words, we want to infer about Y given the data d.

I.1.2. Ground Water Flow

In heterogeneous and fractured media it is essential to understand the vertical distribution of lateral hydraulic conductivity in order to correctly interpret and model groundwater flow and contaminant transport as described in Fienen *et al.* (2004). The objective is to estimate hydraulic conductivity in discrete vertical layers within an aquifer using the flow rate measured with an electromagnetic bore hole flow meter (EBF) positioned, sequentially, at various elevations in the bore hole. The model used to calculate flow (G) given hydraulic conductivity (Y) is called the forward model. The forward model is given by

$$G(Y) = \int_{z_o}^{z_o+h} Y(\xi)d\xi. \tag{1.2}$$

In this equation, the unknown is $Y(\xi)$, the hydraulic conductivity at depth ξ.
$G(Y)$ is the cumulative influx in the interval between the bottom of the bore hole (at elevation z_o) and the elevation of the EBF ($zo + h$). In this example, the forward model can be written as $d = G(Y) + \epsilon$, where d is the observed data observed at different depths and based on that we want to estimate (with uncertainty measures) the hydraulic conductivity Y.

I.1.3. Weather Forecasting

One of the important aspects of weather forecasting is to determine the global velocity field $v(x,t)$ of the air in the atmosphere. Here x denotes the position of the velocity field and t denotes the time $((x,t) \in D \times [0,\infty), D \in R^2)$. The data available are from commercial and military aircraft and weather balloons etc. The main objective here is to find the initial velocity and height fields $(v_0(x), h_0(x)) = (Y(x))$, say. The data we have is noisy observation of velocity field v. For a given Y the forward model of finding the velocity at time t can be solved by a coupled pair of PDEs:

$$\frac{\partial v}{\partial t} = Sv - \nabla h, \quad \frac{\partial h}{\partial t} = -\nabla . v, \tag{1.3}$$

where $S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, ∇ is the differential operator $(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2})$. For a given initial velocity and height Y the velocity field over time can be found by solving (1.3). So we can write $v = G(Y)$, where G is the forward operator. Let us denote the concatenating data as d then the problem can be written as $d = G(Y) + \epsilon$, where G is related to Y by the PDE's (1.3). Here G is called the observational operator.

For the definiteness of the problem the simulated examples and the practical oil-field examples included in this thesis are all from the reservoir characterization example, but it is to be noted that our method can be easily adapted for the other examples.

In the first chapter we consider inverse problems where the input is a high dimension spatial field. The output is the result of a complex system which can be predicted, usually by running a numerical simulator that solves a discretized approximation to a system of non-linear partial differential equations. In addition to the out-put data some data in also available on the spatial field on a coarse-grid. Our goal is to predict the fine-scale spatial field together with the uncertainties in the prediction. We use

Karhunen-Loève (K-L) expansion (see Loève (1977)) of this unknown spatial field. The number of terms in the K-L expansion determines how much information is truly required to capture the variability of the unknown spatial field. We treat this number as an additional model unknown and use the reversible jump Metropolis (RJM) algorithm to handle this random dimension situation. Since the parameters of the covariance function are unknown, at each step of the reversible jump MCMC procedure, we have to use the K-L expansion of the covariance function which is very computationally demanding. Hence, we propose an alternative approach in which we have precomputed the K-L expansion for a given set of the parameters and then use linear interpolation to find the respective eigen pairs for a proposed new value of the parameters. This linear interpolation made the computation much faster. Using the matrix perturbation theory, we have shown that if the interpolating grid is small the approximated eigen values and eigen vectors are very close to the true ones.

We employ a Gaussian process prior for the unknown field and use a hierarchical Bayes model to incorporate multiscale data. In this Bayesian framework, we have shown that this inverse problem is well-posed by proving that the posterior measure is Lipschitz continuous with respect to the data in total variation norm. In our model, the likelihood function contains the forward solver equations (several differential equations) which is not explicitly available and very expensive to compute. Hence, instead of the reversible Jump MCMC algorithm, we propose two-stage reversible jump MCMC. In this algorithm, the proposals are screened in the first stage using the forward solver in a upscaled coarse grid, which is inexpensive due to small dimensions of the coarse grid. Then, it is passed to the final stage only if it has been accepted at the first stage. Thus the two-stage algorithm reduces the computational effort by rejecting the bad proposals at the initial stage. We have shown that this proposed two-stage reversible jump MCMC satisfies the detailed balance condition.

In the second chapter we consider inverse problems in spatial fields with channelized structures. In many geologic environments, the distribution of subsurface properties is primarily controlled by the location and distribution of distinct geologic facies with sharp contrasts in properties across facies boundaries (see Weber (1990)). Under such conditions, the orientation of the channels and channel geometry determine the flow behavior in the subsurface rather than the detailed variations in properties within the channels. Traditional geostatistical techniques for subsurface characterization have typically relied on variograms that are unable to reproduce the channel geometry and the facies architecture (see Haldorsen and Damsleth (1990), Koltermann and Gorelick (1996) and Dubrule (1998)). In this chapter by using hierarchical modeling, the channelized spatial field is represented by facies boundaries and variogram-based spatial fields within each facies. Typically, the parameters representing facies boundaries are highly uncertain, particularly in the early stages of subsurface characterization (see Caumon *et al.* (2004) and Dubrule (1998)). In this chapter the channel boundaries are represented using piecewise linear functions - an approach capable of reproducing a wide variety of channel geometry. The shape of the channel boundaries is updated with dynamic data using reversible jump MCMC where the number of points representing the channel boundaries is assumed to be unknown. To represent variogram-based spatial fields, Karhunen-Loéve expansion (see Loève (1977)) is used. The truncation procedure of the K-L expansion introduce some error in the posterior measure. In this chapter we also estimate a bound for the difference in the expectation of a function with respect to the full and the truncated posterior. Based on this bound, the computation can be simplified by choosing less number of terms in truncation, while the error is in a reasonable range. The sampling of the posterior is done using two stage reversible jump Metropolis-Hastings MCMC.

For the Bayesian inverse problems often the posterior in intractable and we have

to run Markov Chain Monte Carlo algorithms to sample from the posterior. In each iteration of the MCMC we have to run a complex numerical simulator, generally called forward simulator, for the likelihood calculation which is computationally very expensive. In the third chapter we propose to use inexpensive emulator based on a Bayesian approach to multivariate adaptive regression spline (BMARS) for the inverse problems in high dimension spatial field. Difficulty arises while implementing BMARS because here the regressors consists of a high dimension spatial field. So at first we propose to use discrete cosine transformation on the spatial field where the original process is represented by a low dimensional parameterization. The finite dimensional transformed DCT coefficients and the other known input parameters are then used as the regressors while fitting the BMARS for the unknown forward simulator.

CHAPTER II

MULTISCALE DATA INTEGRATION IN LARGE-SCALE SPATIAL INVERSE PROBLEMS

As stated in the introduction, in this chapter we focus on the probelm in which the unknown quantity is a random field (typically signifying process indexed by a spatial coordinate) $Y(x, \omega)$, $x \in D$ and $\omega \in \Omega$ where Ω is a sample space in a probability space (Ω, U, P) with sigma algebra U over Ω, P is a probability measure on U and $D \in R^n$ be a bounded spatial domain. Y is treated as the model variable or the input variable. If Y is known then outcomes (output variables, response) can be predicted, usually by running a numerical simulator that solves a discretized approximation to a system of non-linear partial differential equations. Different names are used in different fields for this model, for example in reservoir simulations the non-linear function that maps the input variables Y to the output $G(Y)$ is called the forward simulator and the concerned modeling problem is called the forward problem. Due to presence of measurement error and other sources of uncertainty, the observed output responses (say d) will be different than that can be produced from this forward model. In an additive model framework, we can relate the observations d to the unknown field Y as

$$d = G(Y) + \epsilon, \tag{2.1}$$

where ϵ is the model error. We assume $\epsilon \sim MVN(0, \sigma_d^2 \mathbf{I})$. The problem we consider here is the inverse of this forward problem, where we want to estimate the model parameter, i.e, the random field Y based on the observations d. A limited number of direct data are available on the spatial field $Y(x, w)$ on a fine grid which will be denoted as y_o. The observed data on the fine-scale spatial field y_o is extremely sparse. Furthermore, additional data on Y may be available on a relatively coarser grid, say

y_c. We like to solve the inverse problem of estimating Y given the data from the output d, the coarse scale data y_c on the spatial field and the data y_o on the fine scale spatial field.

In this chapter we employ a Bayesian Hierarchical model to quantify the uncertainty by formulating the posterior distribution of the fine-scale spatial field $Y(x, w)$ condition on both the coarse-scale data y_c, output data d and the observed fine-scale data y_o. In our model we assume a spatial covariance structure for the fine-scale spatial field and the parameters of the covariance function are updated by the data. To represent variogram-based spatial fields, Karhunen-Loéve expansion (see Loève (1977)) is used. Karhunen-Loéve expansion allows significant reduction in the number of parameters for correlated spatial fields. This is very advantageous in the context of the large scale inverse problem as it allows to perform the search in a smaller parameter space.

In previous findings the number of terms retained in the K-L expansion are taken to be fixed. The criteria used for the number of coefficients retained is monitored by the energy ratio, which was computed before hand. So in those method the number of terms retained in the K-L expansion for each of the proposed spatial may not be enough to capture the true heterogeneity in the model, or may overfit the true spatial field. So we propose to use Reversible Jump MCMC Algorithm where the number of leading terms retained in the K-L Expansion at each of the step is also taken to be random. In other words, in our method the dimension of the parameter space is also random and may change at each step which can be done by Reversible Jump MCMC Algorithm. Our method automatically penalizes for over fitting or under fitting and becomes more flexible to capture the heterogeneity in the spatial field.

Since the parameters of the covariance function is unknown, at each step of the reversible jump MCMC procedure we have to use the K-L expansion of the covariance

function which is very costly in terms of CPU time. So we propose an alternative approach in which we precompute the K-L expansion for a given set of the parameters and then use linear interpolation to find the respective eigen values and eigen vectors for a given new value of the parameters. The linear interpolation is very fast and thus saves us lot of time. Using Matrix perturbation theory we have shown that if the interpolating grid is small the approximated eigen vales and eigen vectors are very close to the true ones. Here our goal is to model the fine-scale spatial field given the water-cut and coarse-scale data, so this is an inverse problem. As we know classical inverse problem is under determined and ill-posed. But here we use a Bayesian framework and we have shown that this Baye's inverse problem is well-posed by proving that the posterior measure is continuous with respect to the data in total variation norm.

As the likelihood function contains the forward solve equations which is very expensive so to save CPU time of the Reversible Jump MCMC iterations we propose to use Two-stage Reversible Jump MCMC where the proposals are screened in the first stage using the forward solve in a upscaled coarse grid, which is inexpensive due to small dimensions of the coarse grid, and is passed to the final stage only if is accepted in the first stage. Thus the two-stage algorithm saves a lot of CPU time by rejecting the bad proposals very fast. We have shown that the Two-stage Reversible Jump MCMC satisfies the Detailed Balance Condition.

In the numerical results, we use priors and hyper priors for the parameters in the covariance functions, error variance, number of terms retained in the K-L expansion. We take the initial spatial to be homogeneous permeabilities, while the reference permeabilities are chosen to be heterogeneous. Our numerical results show that the proposed algorithm can adequately predict the fine scale spatial field. We also observe that if the coarse-scale data is not available we need at least 25% of the fine-scale

data in addition to the output data to predict the fine-scale spatial field adequately. Where as if the coarse-scale data is available which in practical is readily available we observe that with a very few fine-scale measurements in addition to the output data is enough to sample from the fine-scale spatial field efficiently.

The chapter is organized as follows. In the next subsection we discuss the hierarchical Bayes' model and formulate the posterior distribution. In Subsection II.1, Subsection II.2, Subsection II.3 and Subsection II.4 we discus the Bayesian hierarchical model, Metropolis Hastings, reversible jump MCMC and two-stage reversible jump MCMC technique respectively. Finally, in Subsection II.5, we present numerical results.

II.1. Bayesian Framework

We shall explain the general Bayesian framework to solve the inverse problem to infer about the random field Y from the equation

$$d = logit(G(Y)) + \epsilon. \tag{2.2}$$

We have the response data d, some observations on Y at the fine scale denoted by y_o and some coarse scale observations of Y say y_c. The Bayesian solution of the inverse problem will be the posterior distribution of Y conditioned on all the observations which is $P(Y|d, y_c, y_o)$. We express this posterior distribution using the Bayes theorem as

$$P(Y|d, y_c, y_o) \propto P(d|Y, y_c, y_o)P(y_c|Y, y_o)P(y_o|Y)P(Y). \tag{2.3}$$

We need to specify each of the probabilities on the right hand side of the expression to develop the hierarchical Bayesian model. Therefore, the steps to develop the hierarchical Bayes model will be to specify (i) $P(Y)$: the prior model for the un-

known random field Y where we use the Karhunen-Loeve expansion to parameterize Y. (ii) $P(y_o|Y)$: the conditional probability of the fine scale observation given the field Y. $(iii)P(y_c|Y, y_o)$: modeling the coarse scale observation y_c conditioning on the fine scale observations y_o and the Y using the upscaling technique. (iv) $P(d|Y, y_c, y_o)$: the likelihood function which will be obtained from the equation (2.1). In following subsections we provide the details of each of this modeling part.

II.1.1. Modeling the Prior Process $P(Y)$

One of the commonly used stochastic descriptions of spatial fields is based on a two-point correlation function of the spatial field. For spatial fields described with a two-point correlation function, it is assumed that $R(x, y) = E[Y(x, \omega)Y(y, \omega)]$ is known, where $E[\cdot]$ refers to the expectation (i.e., average over all realizations) and x, y are points in the spatial domain. In applications, the spatial fields are considered to be defined on a discrete grid. In this case, $R(x, y)$ is a square matrix with N_{dof} rows and N_{dof} columns, where N_{dof} is the number of grid blocks in the domain. For spatial fields described by a two-point correlation function, one can use the Karhunen-Loève expansion (KLE), following Wong (1971), to obtain spatial field description with possibly fewer degrees of freedom. This is done by representing the spatial field in terms of an optimal L^2 basis. By truncating the expansion, we can represent the spatial matrix by a small number of random parameters.

We briefly recall some properties of the KLE. For simplicity, we assume that $E[Y(x, \omega)] = 0$. Suppose $Y(x, \omega)$ is a second order stochastic process with $E[Y^2(x, \omega] < \infty$, $\forall x \in D$. Given an orthonormal basis $\{\Phi_i\}$ in L^2, we can expand $Y(x, \omega)$ as a general Fourier series $Y(x, \omega) = \sum_i Y_i(\omega)\Phi_i(x)$, where $Y_i(\omega) = \int_D Y(x, \omega)\Phi_i(x)dx$. We are interested in the special L^2 basis $\{\Phi_i\}$ which makes the random variables Y_i uncorrelated. That is, $E(Y_iY_j) = 0$ for all $i \neq j$. The basis functions $\{\Phi_i\}$ satisfy

$E[Y_iY_j] = \int_D \Phi_i(x)dx \int_D R(x,y)\Phi_j(y)dy = 0,\ i \neq j$. Since $\{\Phi_i\}$ is a complete basis in L^2, it follows that $\Phi_i(x)$ are eigenfunctions of $R(x,y)$:

$$\int_D R(x,y)\Phi_i(y)dy = \lambda_i\Phi_i(x), \qquad i = 1,2,\ldots, \tag{2.4}$$

where $\lambda_i = E[Y_i^2] > 0$. Furthermore, we have $R(x,y) = \sum_i \lambda_i\Phi_i(x)\Phi_i(y)$. Denote $\theta_i = Y_i/\sqrt{\lambda_i}$, then θ_i satisfy $E(\theta_i) = 0$ and $E(\theta_i\theta_j) = \delta_{ij}$. It follows that

$$Y(x,\omega) = \sum_i \sqrt{\lambda_i}\theta_i(\omega)\Phi_i(x), \tag{2.5}$$

where Φ_i and λ_i satisfy (2.4). We assume that the eigenvalues λ_i are ordered as $\lambda_1 \geq \lambda_2 \geq \ldots$. The expansion (2.5) is called the Karhunen-Loève expansion. In the KLE (2.5), the L^2 basis functions $\Phi_i(x)$ are deterministic and resolve the spatial dependence of the spatial field. The randomness is represented by the scalar random variables θ_i. After we discretized the domain D by a rectangular mesh, the continuous KLE (2.5) is reduced to finite terms and $\Phi_i(x)$ are discrete fields. Generally, we only need to keep the leading order terms (quantified by the magnitude of λ_i) and still capture most of the energy of the stochastic process $Y(x,\omega)$. For an N_{KL}-term KLE approximation $Y_{N_{KL}} = \sum_{i=1}^{N_{KL}} \sqrt{\lambda_i}\theta_i\Phi_i$, define the energy ratio of the approximation as

$$e(N_{KL}) := \frac{E\|Y_{N_{KL}}\|^2}{E\|Y\|^2} = \frac{\sum_{i=1}^{N_{KL}} \lambda_i}{\sum_{i=1}^{\infty} \lambda_i}. \tag{2.6}$$

If $\lambda_i, i = 1,2,\ldots,$ decay very fast, then the truncated KLE would be a good approximation of the stochastic process in the L^2 sense. There are different type of spatial covariance functions $R(x,y)$ considered in spatial statistics. For example Spherical, Exponential, Squared Exponential or Gaussian and Mattern Class covariance function In our examples, we use squared exponential covariance structure, though the method is not restricted to this particular covariance structure. $R(x,y)$ in this case

is defined as

$$R(x,y) = \sigma^2 \exp\Big(-\frac{|x_1 - y_1|^2}{2l_1^2} - \frac{|x_2 - y_2|^2}{2l_2^2}\Big), \tag{2.7}$$

l_1 and l_2 are the correlation lengths in each direction, and σ^2 is the variance. We reparametrize the spatial field Y by K-L expansion and keep the leading m terms in the KLE. For an m-term KLE approximation

$$Y^m = \theta_0 + \sum_{i=1}^{m} \sqrt{\lambda_i}\theta_i\Phi_i, = B(l_1, l_2, \sigma^2)\theta, (\text{say}), \tag{2.8}$$

where, $B = [\sqrt{\lambda_1}\Phi_1, \sqrt{\lambda_2}\Phi_2 \ldots \sqrt{\lambda_m}\Phi_m]$ and $\theta = (\theta_0, \cdots, \theta_m)$. Here, B depends only on l_1, l_2 and σ^2. Consequently we have a parametric representation of the field Y through $(l_1, l_2, \sigma^2, m, \theta_0, \theta_1, \ldots \theta_m)'$, and we can evaluate Y if we know these parameter values. First, we develop the model and the computation schemes for a fixed m; afterward extend them for unknown m in Subsection II.3. Therefore using Bayes' theorem we can write the posterior $P(Y|d, y_c, y_o)$ as given in equation (2.3) in terms of this set of parameters as

$$\begin{aligned} P(\theta, l_1, l_2, \sigma^2|d, y_c, y_o) &\propto P(d|\theta, l_1, l_2, \sigma^2, y_c, y_o)P(y_c|\theta, l_1, l_2, \sigma^2, y_o) \\ &\quad \times P(y_o|\theta, l_1, l_2, \sigma^2)P(\theta)P(l_1, l_2)P(\sigma^2) \\ &\propto P(d|\theta, l_1, l_2, \sigma^2)P(y_c|\theta, l_1, l_2, \sigma^2)P(y_o|\theta, l_1, l_2, \sigma^2) \\ &\quad \times P(\theta)P(l_1, l_2)P(\sigma^2). \end{aligned} \tag{2.9}$$

II.1.2. Modeling the Fine Scale Data $P(y_o|Y)$

The fine scale observations are obtained at some locations of the field Y and we specify a model $P(y_o|Y)$ or $P(y_o|m, \theta, l_1, l_2)$ as

$$y_o = y_p + \epsilon_k, \tag{2.10}$$

where, y_p is the the fine scale spatial field at the given well locations x^{obs} obtained from the K-L expansion described in Subsection II.1.1. ϵ_k is the model error for the K-L approximation. We assume ϵ_k follows a multivariate normal distribution with mean 0 and covariance $\sigma_k^2 I$. i.e, $y_o|\theta, l_1, l_2, \sigma^2, \sigma_k^2 \sim MVN(y_p, \sigma_k^2)$. The prior for σ_k^2 is assumed to be $\sigma_k^2 \sim InverseGamma(a_k, b_k)$. After integrating out σ_k^2 we obtain

$$P(y_o|\theta, l_1, l_2, \sigma^2) \propto \frac{\Gamma(a_k + N_{obs}/2)}{[b_k + \frac{1}{2}(y_o - y_p)'(y_o - y_p)]^{(a_k + N_{obs}/2)}}, \tag{2.11}$$

where, N_{obs} is he number of observations of the fine-scale permeability field.

II.1.3. Modeling $P(y_c|Y, y_o)$ by Upscaling

In many cases the coarse scale data are readily available which contain important information to reduce the uncertainty in estimation of the fine-scale spatial field. Moreover, solving the forward problem in a coarse-grid is always much faster and we exploit it in our multistage MCMC algorithm. Upscaling procedure is a way to link the coarse and the fine scale data. The simplest way to think about the upscaling procedure in the spatial domain is the use of spatial block averages of the fine-scale spatial data to obtain the coarse scale data. We need to modify this averaging idea in a way so that the forward equations (and the corresponding boundary conditions) remained valid in this upscaling scheme. The main idea of our approach is to upscale the spatial field Y on the coarse-grid, then solve the original system on the coarse-grid with upscaled spatial field (see Christie (1996) Barker and Thibeau (1997) and Durlofsky (1998)). The upscaling procedure depends on the particular forward equations and more details for the permeability field related equation have been provided in Subsection II.5.

The main theme of the procedure is that given a fine scale spatial field Y, we can use a operator L (it can be averaging or more complicated integrations with boundary

conditions) so that the coarse data y_c can be expressed as $y_c = L(Y) + \epsilon_c$, where ϵ_c is a random error term which explains the variations from deterministic upscaling procedures. As we have parameterized the spatial field Y using the K-L expansion the final equation is given as

$$y_c = L(Y) + \epsilon_c = L_c(\theta, l_1, l_2, \sigma^2) + \epsilon_c, \tag{2.12}$$

where L_c can be looked upon as an operator whose input is the fine-scale spatial field or the parameters of the model θ, l_1, l_2 and σ^2 and output is the coarse-scale value at a given location. We assume that the error ϵ_c follows a multivariate normal distribution with mean 0 and covariance $\sigma_c^2 I$. i.e $y_c|\theta, l_1, l_2, \sigma^2, \sigma_c^2 \sim MVN(L_c(\theta, l_1, l_2, \sigma^2), \sigma_c^2 I)$. We assume the prior distribution of σ_c^2 as $\sigma_c^2 \sim InverseGamma(a_c, b_c)$. Furthermore, after integrating out σ_c^2 we obtain the marginal distribution as

$$P(y_c|\theta, l_1, l_2, \sigma^2) \propto \frac{\Gamma(a_c + N^*/2)}{[b_c + \frac{1}{2}||(y_c - L_c(\theta, l_1, l_2, \sigma^2)||^2]^{(a_c + N^*/2)}}. \tag{2.13}$$

where, N^* is he number of observations of the coarse-scale permeability field. The choice of the upscaling operator L_c depends on the forward solver related to the scientific problem. The details about the choice of L_c for the reservoir simulation problem has been provided in Subsection II.5.

II.1.4. Modeling the Likelihood $P(d|Y, y_c, y_o)$

The likelihood is derived from the equation (2.2) as

$$d = G((B(l_1, l_2, \sigma^2)\theta)) + \epsilon_f = F_f(\theta, l_1, l_2, \sigma^2) + \epsilon_f, \tag{2.14}$$

where F_f can be looked upon as a realization from the forward simulator whose input variables are the parameters θ, l_1, l_2 and σ^2. This realization F_f is obtained from the forward simulator through solution of several differential equations. We assume the er-

ror distribution as $\epsilon_f \sim MVN(0, \sigma_f^2 I)$, i.e, $d|\theta, l_1, l_2, \sigma^2, \sigma_f^2 \sim MVN(F_f(\theta, l_1, l_2, \sigma^2), \sigma_f^2 I)$. The prior distribution for σ_f^2 is assumed to be $\sigma_f^2 \sim InverseGamma(a_f, b_f)$. Then, after integrating out σ_f^2, we have the marginal likelihood as

$$P(d|\theta, l_1, l_2, \sigma^2) \propto \frac{\Gamma(a_f + n/2)}{[b_f + \frac{1}{2}||(d - F_f(\theta, l_1, l_2, \sigma^2)||^2]^{(a_f + n/2)}}. \tag{2.15}$$

II.1.5. Prior Distributions

We need to assign prior distribution for the parameters of the covariance kernel. The prior distribution for θ is given by $\theta|\sigma_\theta^2 \sim MVN(0, \sigma_\theta^2 I)$ and $\sigma_\theta^2 \sim InverseGamma(a_0, b_0)$. Again, after integrating out σ_θ^2 we obtain the marginal prior distribution as

$$P(\theta) \propto \frac{\Gamma(a_0 + m/2)}{[b_0 + \frac{1}{2}\theta'\theta]^{(a_0 + m/2)}}. \tag{2.16}$$

Additionally, the prior distribution for σ^2 is taken to be $Gamma(a_s, b_s)$. We assume uniform priors for l_1 and l_2.

II.1.6. The Posterior Distribution and its Continuity

By equation (2.9) we obtain the posterior distribution of the spatial field Y given the output data d, coarse-scale data y_c and the observed fine scale data y_o using the Bayes theorem as:

$$\begin{aligned} P(\theta, l_1, l_2, \sigma^2|d, y_c, y_o) \quad &\propto \quad P(d|\theta, l_1, l_2, \sigma^2) P(y_c|\theta, l_1, l_2, \sigma^2) P(y_o|\theta, l_1, l_2, \sigma^2) \\ &\qquad P(\theta) P(l_1, l_2) P(\sigma^2). \end{aligned} \tag{2.17}$$

Each part of the expressions in the right hand side has been specified in Subsection II.1. For simplicity, all the model unknowns $(l_1, l_2, \sigma^2, \theta_o, \theta_1 \ldots \theta_m)'$ are denoted as τ in further discussions. Following similar notations $F_f(\theta, l_1, l_2, \sigma^2)$ is denoted by F_τ and $L_c(\theta, l_1, l_2, \sigma^2)$ is denoted by L_τ.

Using (4.20), (2.13),(2.15),(4.15), (2.17) the posterior distribution is given by

$$\begin{aligned}\pi(\tau) = P(\tau|d, y_c, y_o) &\propto \frac{1}{[b_f + \frac{1}{2}||d - F_\tau||^2]^{(a_f+n/2)}} \times \frac{1}{[b_c + \frac{1}{2}||y_c - L_\tau||^2]^{(a_c+N^*/2)}} \\ &\times \frac{1}{[b_k + \frac{1}{2}||y_o - y_p||^2]^{(a_k+N_{obs}/2)}} \times \frac{1}{[b_0 + \frac{1}{2}\theta'\theta]^{(a_0+m/2)}} \\ &\times (\sigma^2)^{a_s-1} \exp(\sigma^2/b_s), \end{aligned} \tag{2.18}$$

where $\|d - F_\tau\|^2 = \sum_{i=1}^{n}(d_i - F_\tau(i))^2$ for n output observations.

We all know that this inverse problem without the proper regularization is under determined and ill-posed. In the Bayesian framework, if one can show that the posterior measure is Lipschitz continuous with respect to the data in the total variation distance, then it guaranties that this Bayesian inverse problem is well-posed (Cotter *et al.* (2009)). To prove it, we assume that the spatial field is given on a finite grid. This assumption is practical because usually spatial field is not defined on very small scales (e.g., pore scale). To show the continuity of the posterior with respect to data, we define

$$\pi_z(\tau) = \frac{1}{Z} g(\tau, z)\pi_0(\tau), \tag{2.19}$$

where z is the concatenating dataset, i.e, $z = \begin{pmatrix} d \\ y_c \\ y_o \end{pmatrix}$.

Furthermore,

$$\begin{aligned} g(\tau, z) = \frac{1}{[b_f + \frac{1}{2}||d - F_\tau||^2]^{(a_f+n/2)}} &\times \frac{1}{[b_c + \frac{1}{2}||y_c - L_\tau||^2]^{(a_c+N^*/2)}} \\ &\times \frac{1}{[b_k + \frac{1}{2}||y_o - y_p||^2]^{(a_k+N_{obs}/2)}}, \end{aligned} \tag{2.20}$$

$$\pi_0(\tau) = \frac{1}{[b_0 + \frac{1}{2}\theta'\theta]^{(a_0+m/2)}} \times (\sigma^2)^{a_s-1} \exp(\sigma^2/b_s), \tag{2.21}$$

$$Z = \int g(\tau, z)\pi_0(\tau)d\tau. \tag{2.22}$$

Theorem II.1.1. *$\forall\ r > 0,\ \exists\ C = C(r)$ such that the posterior measures π_1 and π_2 for two different data sets z_1 and z_2 with $max\left(\|z_1\|_2, \|z_2\|_2\right) \leq r$, satisfy*

$$\|\pi_1 - \pi_2\|_{TV} = \frac{1}{2}\int \left|Z_1^{-1} g(\tau, z_1) - Z_2^{-1} g(\tau, z_2)\right| \pi_0(\tau) d\tau \leq C\|z_1 - z_2\|_2, \qquad (2.23)$$

where Z_1 and Z_2 are defined by (2.22) for z_1 and z_2, respectively.

The proof is given in the Appendix A. Note that it can also be shown that the above Lipschitz continuity condition is also valid for Hellinger distance, i.e

$$\begin{aligned} d_{Hell}(\pi_1 - \pi_2) &= \frac{1}{2}\left(\int \left(\sqrt{Z_1^{-1} g(\tau, z_1)} - \sqrt{Z_2^{-1} g(\tau, z_2)}\right)^2 \pi_0(\tau) d\tau\right)^{-1/2} \\ &\leq C\|z_1 - z_2\|_2. \end{aligned} \qquad (2.24)$$

II.2. Bayesian Computation Using MCMC

As the posterior is not analytically tractable, hence we use MCMC based computation method to simulate the parameters from the posterior distribution. First, we consider the case where we fix the number of terms retained in the K-L expansion. We solve the eigen value problem for the fine-scale spatial field beforehand and select an m, number of terms retained in K-L expansion, such that the energy ratio defined in (2.6) is at least 90%. For a constant m we use the standard Metropolis Hastings MCMC to sample from the posterior.

Algorithm (Metropolis-Hastings MCMC), Robert and Casella (1999)
Suppose at the r^{th} step we are at the state τ_r, then

- Step 1. Generate τ^* from $q(\tau^*|\tau_r)$.

- Step 2. Accept τ^* with probability

$$\alpha(\tau_r, \tau^*) = \min\left\{1, \underbrace{\frac{P(d|\tau^*)P(y_c|\tau^*)P(y_o|\tau^*)}{P(d|\tau_r)P(y_c|\tau_r)P(y_o|\tau_r)}}_{\text{likelihood ratio}} \times \underbrace{\frac{P(\tau)}{P(\tau_r)}}_{\text{prior ratio}} \times \underbrace{\frac{q_{\tau^*}(\tau_r|\tau^*)}{q_{\tau^*}(\tau^*|\tau_r)}}_{\text{proposal ratio}}\right\} \quad (2.25)$$

  i.e $\tau_{r+1} = \tau^*$ with probability $\alpha(\tau_r, \tau^*)$, and $\tau_{r+1} = \tau_r$ with probability $1 - \alpha(\tau_r, \tau^*)$.

Starting with an initial parameters of the spatial sample τ_0, the MCMC algorithm generates a Markov chain $\{\tau_r\}$. The target distribution $\pi(\tau)$ is the stationary distribution of the Markov chain τ_r, so τ_r represent the samples generated from $\pi(\tau)$ after the chain converges and reaches a steady state. As an example we can use standard random walk Metropolis–Hastings algorithm to generate samples from the posterior distribution. Then at the r^{th} step, we propose $\tau^* = \tau_r + h_\tau u_\tau$, where u_τ is generated from a $N(0, I)$ distribution.

Here at each iteration step after we propose a new $\theta, l_1, l_2, \sigma^2$, we have to solve the eigen value problem for the K-L expansion to get the fine-scale spatial realizations which is very expensive. To speed up the computation, we compute the eigen value problem (K-L expansion) for a certain number of pairs of l_1, l_2 beforehand and interpolate them to find the eigen values and eigen vectors at each step in the Metropolis Hastings MCMC. Note that change of σ doesn't change the eigen vectors, it only changes the magnitude of the eigen values which can be adjusted by a scale factor. We can show that this approximation is valid if the interpolation grid of the correlation length is sufficiently small. Since the magnitude of sigma doesn't effect the interpolation without loss of generality we can assume $\sigma^2 = 1$. Also here we prove only the isotropic case i.e $l_1 = l_2 = l$, (say).

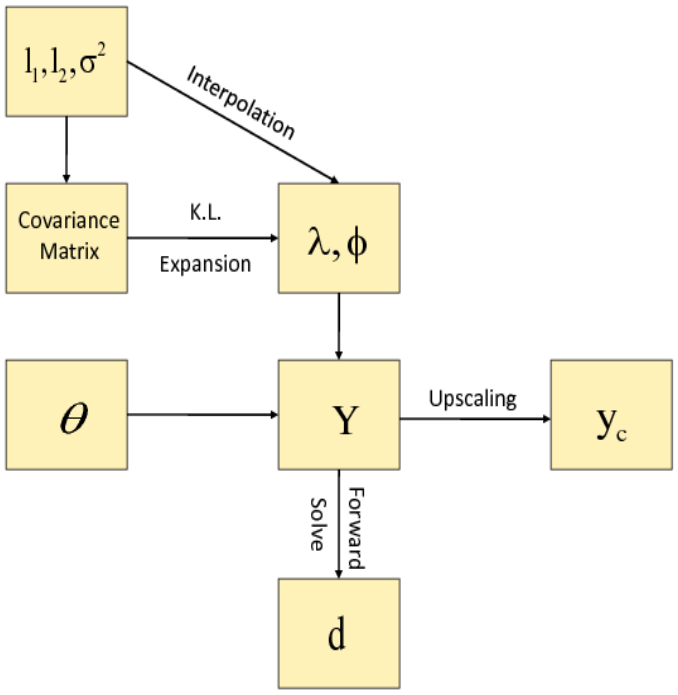

Fig. 2. Flow-chart for the hierarchical model.

Theorem II.2.1. *Suppose A_l be the covariance matrix for a given correlation length, l. Let $\lambda_1, \lambda_2 \ldots, \lambda_m$ be the m ordered eigen values considered in the K-L expansion of A_l and let $\phi_1, \phi_2 \ldots, \phi_m$ be the corresponding orthonormal eigen vectors. Suppose $A_{l+\delta l}$ be the covariance matrix if we perturb the correlation length l by a small quantity δl. Let $\lambda_1^{'}, \lambda_2^{'} \ldots, \lambda_m^{'}$ be the m ordered eigen values considered in the K-L expansion and let $\phi_1^{'}, \phi_2^{'} \ldots, \phi_m^{'}$ be the corresponding orthonormal eigen vectors. then,*

$$\lambda_i^{'} = \lambda_i + O(\delta l), \text{ and } \phi_i^{'} = \phi_i + O(\delta l), \ \forall i. \tag{2.26}$$

The proof is given in Appendix B using matrix perturbation Theory.

The hierarchical model and the MCMC procedure is described by a simple graphical model on Figure 2.

II.3. Extension to Model with Unknown m

In the previous analysis, m, the dimension of θ remained fixed, so the number of the terms retained in K-L expansion is taken to be a constant. Usually it has been estimated by using equation (2.6). This method only utilizes the fine scale direct data y_o but ignores the output data d and the coarse scale data y_c. That way this approach may not capture the actual heterogeneity of the spatial field very well. We extend our previous model by treating m as an additional model unknown and obtain its posterior distribution by conditioning on all the available data. In this situation, we develop the Bayesian hierarchical model by extending equation (2.17) as

$$\begin{aligned} P(m,\theta,l_1,l_2,\sigma^2|d,y_c,y_o) &\propto P(d|m,\theta,l_1,l_2,\sigma^2)P(y_c|m,\theta,l_1,l_2,\sigma^2) \\ &\times P(y_o|m,\theta,l_1,l_2,\sigma^2)P(\theta|m)p(m)P(l_1,l_2)P(\sigma^2) \end{aligned} \tag{2.27}$$

We keep all the model specifications same as in Subsection II.1 but use a truncated Poisson prior for $P(m)$. We need to modify the MCMC computation procedure due to this unknown dimension. If we vary the number of terms in K-L expansion then the dimension of θ will also change in each step. This jumping between different dimensions in the parameter space can be achieved through reversible jump Markov Chain Monte Carlo methods as proposed by Green (1995). We describe the reversible jump MCMC procedure in our case following the general approach of the reversible jump MCMC (Waagepetersen and Sorensen (2001)).

In our case using the hierarchical Bayes' model, the posterior can be written as

$$\begin{aligned} \pi(\tau,m) &\propto P(d|\theta,l_1,l_2,\sigma^2,m)P(y_c|\theta,l_1,l_2,\sigma^2,m) \\ &\times P(y_o|\theta,l_1,l_2,\sigma^2,m)P(l_1,l_2)P(\sigma^2)P(\theta|m)P(m). \end{aligned} \tag{2.28}$$

We assume the prior for $m|\lambda$ as truncated Poisson(λ), truncated at m_{max}, where

$\lambda \sim Gamma(\nu, \beta)$. Integrating out λ we get $P(m) \propto \frac{1}{(1/\beta+1)^{(m+\nu+1)}}$. All the other terms remains same as in (2.18).

Algorithm: Reversible Jump MCMC as Birth and Death Process.

Suppose at the r^{th} step we are at the state (m_r, τ_r), then we have three possible steps:

- Birth Step. Propose to add the $(m_r + 1)^{th}$ term in the K-L expansion with probability $p^b_{m_r}$. Propose θ' from $q(.)$ and hence $\theta^* = (\theta_r, \theta')$. The acceptance probability is given by $\alpha_{m_r, m_r+1}(\theta_r, \theta^*) = min\{1, \frac{\pi(\theta^*, m_r+1)p^d_{m_r+1}}{\pi(\theta_r, m_r)p^b_{m_r} q(\theta')}\}$.

- Death Step. Propose to delete the $(m_r)^{th}$ term with probability $p^d_{m_r}$. So here $(\theta^*, \theta_r^{m_r}) = \theta_r$. The acceptance probability is given by $\alpha_{m_r, m_r-1}(\theta_r, \theta^*) = min\{1, \frac{\pi(\theta^*, m_r-1)p^b_{m_r-1} q(\theta_r^{m_r})}{\pi(\theta_r, m_r)p^d_{m_r}}\}$.

- Jump Step. Propose a new θ with the same dimension along with l_1, l_2, σ^2 with probability $p^s_{m_r}$. In other words generate τ^* from $q(\tau^*|\tau_r)$. The acceptance probability is given by $\alpha(\tau_r, \tau^*) = min\{1, \frac{\pi(\tau^*)q(\tau_r|\theta^*)}{\pi(\tau_r)q(\tau^*|\tau_r)})\}$.

Where, $p^b_{m_r} + p^d_{m_r} + p^s_{m_r} = 1, \forall m_r$.

II.4. Two Stage Reversible Jump MCMC

The main disadvantage of the above reversible jump MCMC algorithm is the high computational cost in solving the forward model on the fine-grid to compute G in the target distribution $\pi(\tau, m)$. Typically, in our simulations, reversible jump MCMC method converges to the steady state after several iterations. That way, a large amount of CPU time is spent on simulating the rejected samples, making the direct (full) reversible jump MCMC simulations very expensive.

The direct reversible jump MCMC method can be improved by adapting the proposal distribution $q(\tau, m|\tau_n, m_n)$ to the target distribution using a coarse-scale

model. This can be achieved by a two-stage reversible jump MCMC method, where we compare the output from the forward model on a coarse-grid , first. If the proposal is accepted by the coarse-scale test, then a full fine-scale computation will be conducted and the proposal will be further tested as in the direct reversible jump MCMC method. Otherwise, the proposal will be rejected by the coarse-scale test and a new proposal will be generated from $q(\tau, m|\tau_n, m_n)$. The coarse-scale test filters the unacceptable proposals and avoids the expensive fine-scale tests for those proposals. The filtering process essentially modifies the proposal distribution $q(\tau, m|\tau_n, m_n)$ by incorporating the coarse-scale information of the problem. The algorithm for a general two-stage MCMC method with upscaling was introduced in Efendiev *et al.* (2007). Our hierarchical model can also take an advantage of inexpensive upscaled simulations to screen the proposals. Here we extend the algorithm to two-stage reversible jump MCMC method. Let F_τ^* be the output computed by solving the forward model on a coarse-scale for the given fine-scale spatial field with parameters (τ, m). In the case of Reservoir characterization this is done either with upscaling methods or mixed MsFEM. The fine-scale target distribution $\pi(\tau, m)$ is approximated on the coarse scale by $\pi^*(\tau, m)$. Here all the terms in the expression of $\pi^*(\tau, m)$ is same as that of $\pi(\tau, m)$ except only the likelihood term $\frac{1}{[b_f+\frac{1}{2}||d-F_\tau||^2]^{(a_f+n/2)}}$ is replaced by $\frac{1}{[b_f+\frac{1}{2}H||d-F_\tau^*||^2]^{(a_f+n/2)}}$. Where the function H is estimated based on offline computations using independent samples from the prior. More precisely using independent samples from the prior distribution, the spatial fields are generated. Then both the coarse-scale and fine-scale simulations are performed and $\|d - F_\tau\|$ vs $\|d - F_\tau^*\|$ are plotted. This scatter plot data can be modeled by $\|d - F_\tau\| = H(\|d - F_\tau^*\|) + w$, where w is a random component representing the deviations of the true fine-scale error from the predicted error. Using the coarse-scale distribution $\pi^*(\tau)$ as a filter, the two-stage reversible jump MCMC can be described as follows.

Algorithm: Two-stage reversible jump MCMC as Birth and Death Process.

Suppose at the n^{th} step we are at the state ν_n. Let k_n be the corresponding fine-scale permeability field. Here $\nu_n = (\tau_n, m_n)$.

- Step 1. This step is the same as the reversible jump MCMC method described earlier. The only difference is the fractional flow F_ν^* is computed by solving the coarse-scale model. At ν_n, generate a trial proposal $\tilde{\nu}$ from distribution $q(\tilde{\nu}|\nu_n)$ the same way as in the reversible jump MCMC described earlier i.e this step is same as doing reversible jump MCMC on $\pi^*(\nu)$.

- Step 2. Take the proposal as

$$\nu = \begin{cases} \tilde{\nu} & \text{with probability } \alpha_p(\nu_n, \tilde{\nu}), \\ \nu_n & \text{with probability } 1 - \alpha_p(\nu_n, \tilde{\nu}). \end{cases}$$

  If we are at Birth Step then the acceptance probability is given by

$$\alpha_p(\nu_n, \tilde{\nu}) = min\{1, \frac{\pi^*(\tilde{\tau}, m_n + 1) p^d_{m_n+1}}{\pi^*(\tau_n, m_n) p^b_{m_n} q(\theta')}\}. \tag{2.29}$$

  If we are at death step then the acceptance probability is given by

$$\alpha_p(\nu_n, \tilde{\nu}) = min\{1, \frac{\pi^*(\tilde{\tau}, m_n - 1) p^b_{m_n-1} q(\theta_n^{m_n})}{\pi^*(\tau_n, m_n) p^d_{m_n}}\}. \tag{2.30}$$

  If we are going to have only jumps then the acceptance probability is given by

$$\alpha_p(\nu_n, \tilde{\nu}) = min\{1, \frac{\pi^*(\tilde{\tau}, m_n) q(\tau_n|\tilde{\tau})}{\pi^*(\tau_n, m_n) q(\tilde{\tau}|\tau_n)})\}. \tag{2.31}$$

- Step 3. Accept ν as a sample with probability

$$\alpha_f(\nu_n, \nu) = \min\left(1, \frac{Q(\nu_n|\nu)\pi(\nu)}{Q(\nu|\nu_n)\pi(\nu_n)}\right). \tag{2.32}$$

Where, $Q(\nu|\nu_n)$ is the transition kernel of the first stage. The acceptance probability (2.32) can be simplified as

$$\alpha_f(\nu_n, \nu) = \min\left(1, \frac{\pi(\nu)\pi^*(\nu_n)}{\pi(\nu_n)\pi^*(\nu)}\right). \tag{2.33}$$

To show that the two stage reversible jump MCMC sampling generates a Markov chain, whose stationary distribution is the candidate distribution it is sufficient to show that the transition kernel satisfies the detailed balance condition.

Theorem II.4.1. *If $K(\nu_n, \nu)$ is transition kernel of the Markov Chain ν_n generated by the two-stage reversible jump MCMC, then*

$$\pi(\nu_n)K(\nu_n, \nu) = \pi(\nu)K(\nu, \nu_n). \tag{2.34}$$

The proof is given in the Appendix C.

II.5. Simulated and Real Examples from Reservoir Model

For definiteness our examples are focused on petroleum reservoir models. However the developed theory, methodology and the computation tools will be useful for any inverse problem in spatial and temporal field with large scale simulator. Reservoir simulation models are widely used by oil and gas companies for production forecasts and for making investment decisions. If it were possible for Geo scientists and engineers to know the physical properties like locations of oil and gas, the permeability, the porosity, and the multi-phase flow properties at all locations in a reservoir, it would be conceptually possible to develop a mathematical model that could be used to predict the outcome of any action. This model is usually a set of partial differential equations. If the model variables are known, outcomes (output variables) can be predicted, usually running a numerical reservoir simulator that solves a dis-

cretized approximation to those partial differential equations. This is the forward problem. Unfortunately, most oil and gas reservoirs are inconveniently buried beneath thousands of feet of overburden. Direct observations of physical properties of the reservoir are available only at a few well locations. Additionally, we have some indirect observations known as the production data (the output data d) which are typically made at the surface, either at the well-head or at distributed locations. The main intention is to determine the plausible physical properties of the reservoir given these direct and indirect observations. This is an inverse problem and the solution of the inverse problem provides an estimate of the characteristics of the subsurface media. In order to solve this inverse problem, the mismatch between simulated (from the numerical reservoir simulator) and observed measurements of production data is minimized. This method is known as the history matching in petroleum engineering. The characteristics of the subsurface media are quantified by several parameters such as permeability, porosity, fluid saturation etc., which are the major contributors to the uncertainties in reservoir performance forecasting. Large uncertainties in reservoirs can greatly affect the production and decision making on well drilling. Better decisions can be made by reducing the uncertainty. Thus, quantifying and reducing the uncertainty are important and challenging problems in subsurface modeling. In the following examples, we are particularly interested in quantifying and reducing the uncertainties for one of the major characteristics of subsurface property, permeability. Hence we need to infer about the permeability spatial filed based on the direct and indirect observations. The fine-scale data represent point measurements such as well logs and cores where as the coarse-scale data are obtained from seismic traces. First, we perform simulation studies to explore the behavior of our modeling approach.

II.5.1. The Mathematical Model and Specification of G

The model has been described in Subsection I.1.1 as $d = logit[G(Y)] + \epsilon$. Where d is the watercut data, Y is the fine-scale permeability field expressed in a logarithm scale, i.e $Y = log(k_f)$ and G is simulator output by using the log-permeability field Y. G is determined by the Darcy's law and given by the following PDEs. We consider two-phase flow in a subsurface formation over a bounded set $D \subset R^2$ under the assumption that the fluid displacement is dominated by viscous effects. For clarity of exposition, we neglect the effects of gravity, compressibility, and capillary pressure, although our proposed approach is independent of the choice of physical mechanisms. Also, porosity ϕ will be considered to be known constant. The two phases will be referred as the water and the oil (or a non-aqueous phase liquid), designated by subscripts w and o, respectively. Darcy's law for each phase j, $(j = w, o)$ is given by

$$v_j = -\frac{k_{rj}(S)}{\mu_j} k_f \nabla p, \tag{2.35}$$

where, $v_j = v_j(x,t)$, $(x,t) \in D \times [0,\infty)$, is the phase velocity at time t and spatial location x. μ_j is the dynamic viscosity and $k_f = k_f(x), x \in D$, is the fine-scale permeability field. $S = S(x,t)$, $(x,t) \in D \times [0,\infty)$, is the water saturation (volume fraction) and $k_{rj} = k_{rj}(S(x,t))$, $(x,t) \in D \times [0,\infty)$ is the relative permeability to phase j $(j = o, w)$. For simplicity, here we take $k_{rw} = S^2$ and $k_{ro} = (1-S)^2$. $p = p(x,t)$, $(x,t) \in D \times [0,\infty)$, is the pressure and $\nabla$is the differential operator $\left(\frac{\partial(.)}{\partial x_1}, \frac{\partial(.)}{\partial x_2}\right)$. Combining the Darcy's law with a statement of conservation of mass allows us to express the governing equations in terms of pressure and saturation equations:

$$\nabla \cdot (\lambda(S) k_f \nabla p) = q_1, \tag{2.36}$$

$$v = v_w + v_o = -\lambda(S) k_f \cdot \nabla p, \tag{2.37}$$

$$\phi \frac{\partial S}{\partial t} + v \cdot \nabla f(S) = -q_2, \tag{2.38}$$

where, λ is the total mobility $\lambda(S) = \frac{k_{rw}(S)}{\mu_w} + \frac{k_{ro}(S)}{\mu_o}$, f is the fractional flux of water , $f(S) = \frac{k_{rw}(S)/\mu_w}{k_{rw}(S)/\mu_w + k_{ro}(S)/\mu_o}$, v is the total velocity, q_1 and q_2 are known source term (for simplicity we assume $q_1 = 0$). The above descriptions are referred to as the fine-scale model of the two-phase flow problem. The p.d.e's (2.36)-(2.38), are solved with the following boundary conditions. The pressure p is taken to be known at the boundaries of D, i.e $p = q_3$ for $(x,t) \in \partial D \times [0,\infty)$, where q_3 is known. The initial conditions for S are specified for $t = 0$. We partition ∂D into three parts ∂D^{in}, ∂D^{out} and ∂D^{other}, i.e, $\partial D = \partial D^{in} \cup \partial D^{out} \cup \partial D^{other}$. Where, water in injected on ∂D^{in} and the fluid (oil and water) are produced on ∂D^{out}, also called the production edge. So the boundary condition on S is taken as $S = 1$ on ∂D^{in}. Solving the p.d.e's (2.36)-(2.38) with the above boundary condition for a known permeability field $k_f(x)$ would yield the solution of $v(x,t)$ and $S(x,t)$ $\forall (x,t)$. The fractional flow or the water-cut at time t is defined as the the fraction of water in the produced fluid on the production edge ∂D^{out} and is given by:

$$F(t) = \frac{\int_{\partial D^{out}} v_n f(S(x,t))dx}{\int_{\partial D^{out}} v_n dx}, \tag{2.39}$$

where, v_n is the component of the velocity field v which is normal to the boundary ∂D^{out} and dx denotes that the integration is taken along the boundary. The fractional flow or water-cut at time t, $F(t)$ depends on the total velocity v and the water saturation S, which are the solutions of the pde's (2.36)-(2.38) for a given spatial permeability field $k_f(x) = exp(Y(x))$ with some boundary conditions on S and p. In other words $Y(x)$ is the input and $F(t)$ is the output for the forward simulator. So $F(t)$ can be written as $F(t) = G(Y(x))$. Since $F(t)$ is always between 0 and 1 we

take a logit transformation on G and write the forward model as:

$$d = logit(G(Y(x))) + \epsilon. \tag{2.40}$$

II.5.2. The Upscaling Procedure

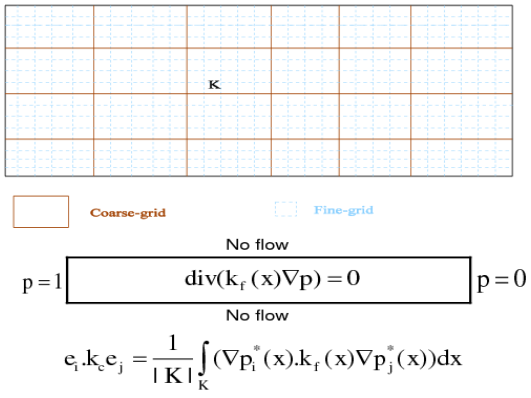

Fig. 3. Schematic description of fine- and coarse-grids. Bold lines illustrate a coarse-scale partitioning, while thin lines show a fine-scale partitioning within coarse–grid cells.

Consider the fine-scale spatial field which is defined in the domain with underlying fine grid as shown in Figure 3. On the same graph we illustrate a coarse-scale partition of the domain. Here we consider single-phase flow upscaling procedure for two-phase flow in heterogeneous porous media (see Efendiev *et al.* (2005)). To calculate the upscaled spatial field at the coarse-level, we use the solutions of local pressure equations. The main idea of the calculation of a coarse-scale permeability is that it delivers the same average response of the forward model as that of the underlying fine-scale problem locally in each coarse-block. For each coarse domain K, we solve

the local pressure equations in the fine grid

$$v = -k_f(x)\nabla p,\ \nabla.(v) = 0, \tag{2.41}$$

with some coarse-scale boundary conditions. Here $k_f(x)$ denotes the fine-scale permeability field, p is the pressure, v is the velocity and ∇ is the partial differential operator $\left(\frac{\partial(.)}{\partial x_1}, \frac{\partial(.)}{\partial x_2}\right)$. The pressure equations in (2.41) can be also written together as:

$$div(k_f(x)\nabla p) = 0. \tag{2.42}$$

The approach considered here is to replace k_f with upscaled coarse permeabilities, k_c, which is constant in each fine grid within the same coarse block. By definition k_c is a discrete quantity relying on the discretization of the medium. In particular, k_c depends on the location and geometry of the grid-block in which it is computed. The essential requirement of k_c is that it leads to pressure and velocity solutions with desired accuracy so that the average response of the forward model in each coarse domain is almost same as the response from the fine-scale model. k_c is defined in a given coarse block K such that

$$k_c \left\langle \nabla p^* \right\rangle_K = - \left\langle v^* \right\rangle_K , \tag{2.43}$$

where p^* and v^* are the solutions of (2.41) (with appropriate boundary conditions, discussed below), and $\langle . \rangle_K = \frac{1}{|K|} \int_K (.)dx$ is the block average over K. In two dimensional setting we solve two independent local fine scale solutions v_i^* and p_i^* of the pressure equations

$$v_i = -k_f(x)\nabla p_i,\ \nabla.(v_i) = 0. \tag{2.44}$$

in each coarse blocks K, $i = 1, 2$, with a given boundary condition. Let e_i be the unit vector in the ith direction. A typical boundary condition is given by $p_i = 1$ and

$p_i = 0$ on the opposite sides along the direction e_i and no flow boundary conditions on all other sides, i.e $v_i.n = 0$, along the direction e_j, $(j \neq i)$ (as shown in Figure 3), where, $v_i.n$ is the component of the velocity field v_i which is normal to the boundary ∂D. For these boundary conditions, the coarse-scale permeability tensor is given by

$$e_i.k_c e_j = \frac{1}{|K|} \int_K (\nabla p_i^*(x).k_f(x) \nabla p_j^*(x)) dx, \tag{2.45}$$

where p_i^* and p_j^* are the solution of (2.41) with corresponding boundary conditions. Various boundary condition can have some influence on the accuracy of the calculations, including periodic, Dirichlet etc. (see Wu *et al.* (2002)). In our numerical examples we take the logarithm of the observed coarse scale permeability as our coarse data, i.e $y_c = log(k_c)$.

II.5.3. Numerical Results for Simulated Reservoirs

In our first example, we have considered only the isotropic case, i.e we take $l_1 = l_2 = l$, (say). We consider a 50×50 fine-scale permeability field on the unit square. We generate 15 fine-scale permeability field with $l = .25$, $\sigma^2 = 1$ and the reference permeability field is taken to be the average of these 15 permeability field. The fractional flow or water cut data is generated from the reference permeability field using eclipse software and was validated by the petroleum engineering department. The observed coarse-scale permeability field is calculated using the upscaling procedure in a 5×5 coarse grid. The fine-scale permeability field is observed at 6 locations along $x = 0$ and $x = 1$ boundaries. First we implement the reversible jump MCMC algorithm. We take the first 20 terms in the K-L expansion while generating the reference field. The mode of the posterior distribution of m comes out to be 19. The posterior mean of fine-scale permeability field resembles very close to the reference permeability field. The mode of the posterior density of l is near 0.25. The posterior density σ^2 are

centered around 1. So, the posterior density of σ^2 and l have peak corresponding to the values of the generated reference field. Then we implement two-stage reversible jump MCMC algorithm with the same reference field and water cut data as we have used in the reversible jump MCMC. The two-stage reversible jump MCMC produced the same results as in reversible jump MCMC (see Figures 4, 5 and 6). The two-stage algorithm is much faster as it rejects the bad samples in the first-stage where we solve the partial differential equations on a coarse grid. The effective acceptance rate of the two-stage algorithm increases to almost eighty percent where as the regular Reversible Jump MCMC have a acceptance rate of nearly ten percent. Then we consider a case where we assume that no coarse-scale data is available but ten percent of the fine scale data are available at equidistant point. We proceed with the same reference permeability field and same water-cut data. The results are plotted on Figure 7. The same procedure is replicated assuming twenty five percent data available (see Figure 8). We can see that if coarse-scale data is not available ten percent fine scale data is not enough to capture the parameters of the model, we need at least twenty-five percent of the fine-scale data.

II.5.4. Numerical Results for a Real Field Example

In this subsection we apply our model on a real field example , viz. punq-s3 model dataset. The PUNQ-S3 case has been taken from a reservoir engineering study on a real field performed Elf Exploration Production. It was qualified as a small-size industrial reservoir engineering model. The model contains $19 \times 28 \times 5$ grid blocks. The PUNQ-S3 data set was an experimental study where the true permeability was actually known on the $19 \times 28 \times 5$ grid but the researchers were asked not to use the permeability data for their modeling purpose. They were asked to use the production history only to infer about the true permeability field and then compare how their

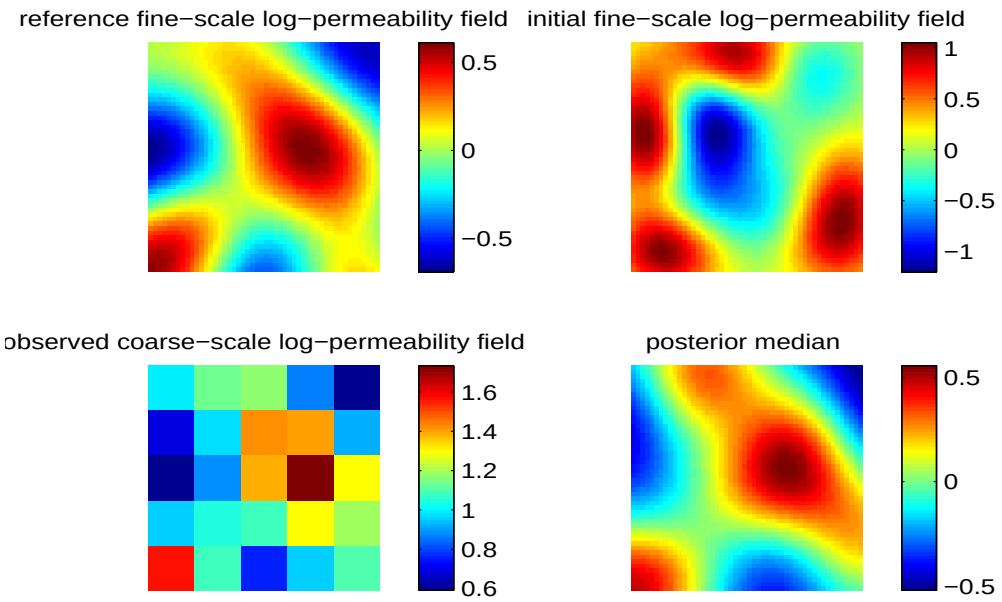

Fig. 4. Log permeability plot for the simulated example using two stage reversible jump MCMC. Top left: The true fine-scale log permeability field. Top right: Initial fine-scale log permeability field. Bottom left: The observed coarse-scale permeability field. Bottom Right: The median of the sampled fine-scale permeability field.

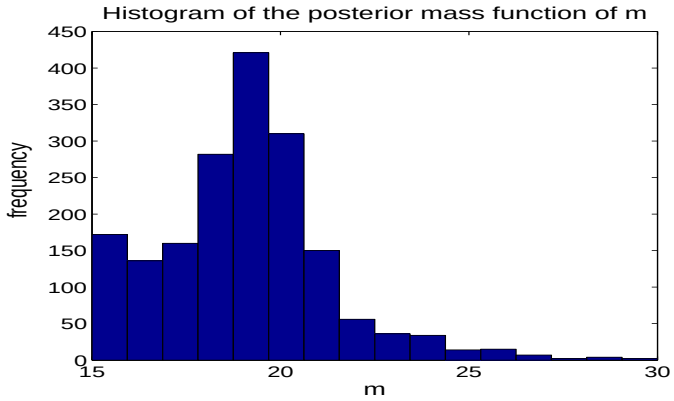

Fig. 5. Histogram of the posterior distribution of m for two-stage reversible jump MCMC.

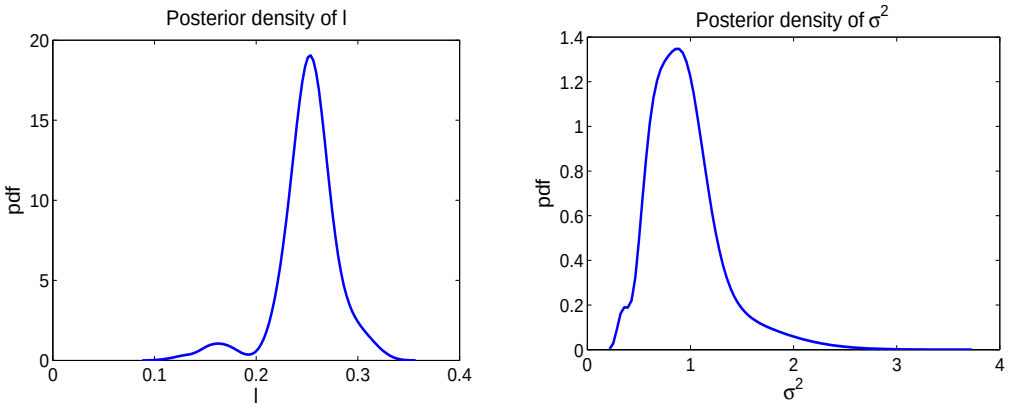

Fig. 6. Posterior densities for the simulated example using two-stage reversible jump MCMC. Left: Posterior density of l, Right: Posterior density of σ^2.

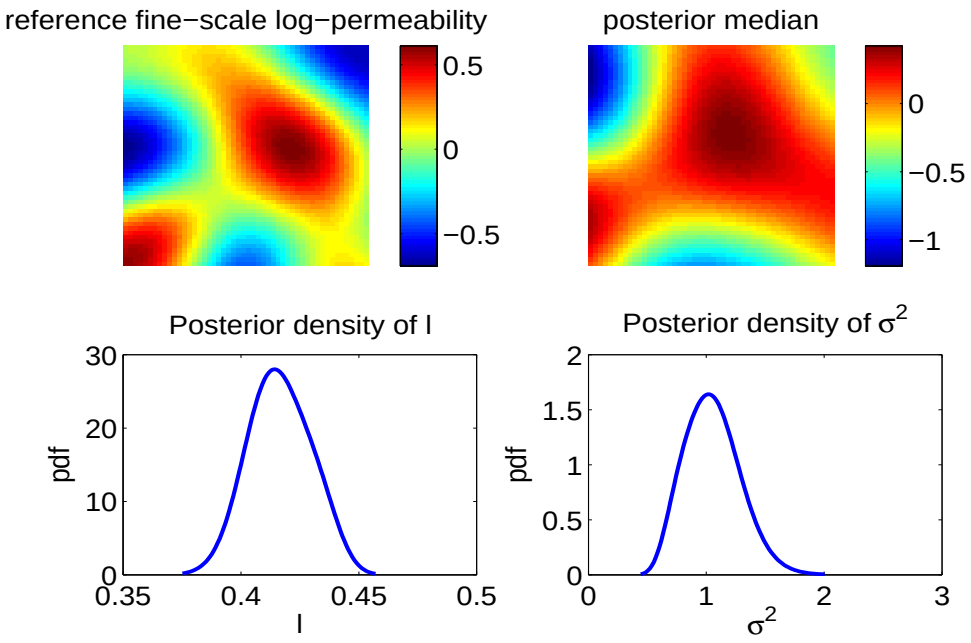

Fig. 7. Log permeability plot with only 10 percent fine-scale data observed and no coarse-scale data available. The true fine-scale log permeability field. Top right: The median of the sampled fine-scale permeability field. Bottom left: Posterior density of l, Bottom Right: Posterior density of σ^2.

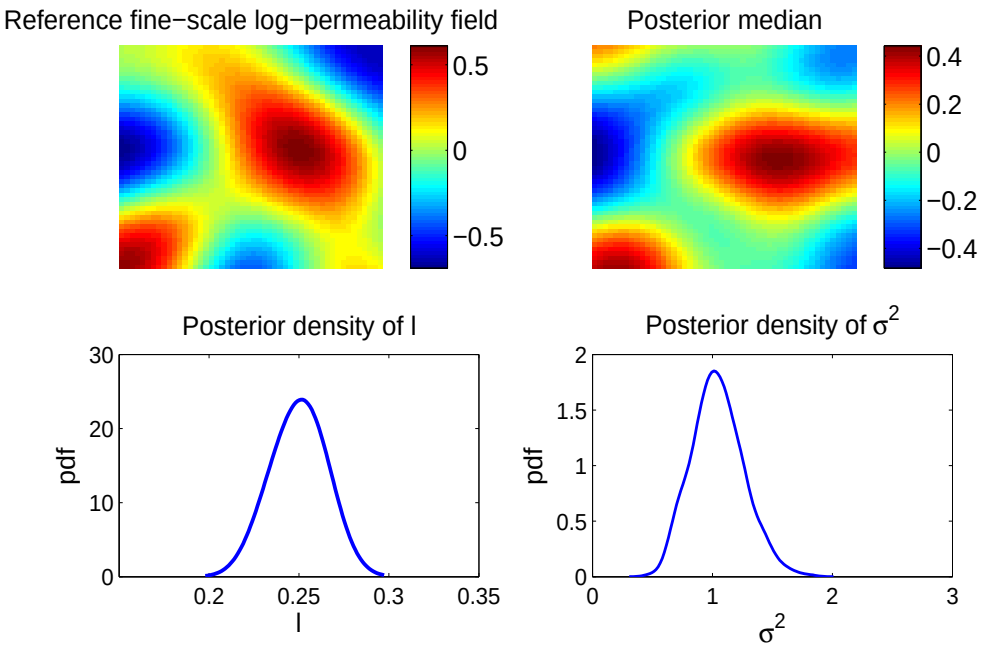

Fig. 8. Posterior densities with only 25 percent fine-scale data observed and no coarse-scale data available. Top left: The true fine-scale log permeability field, Top right: The median of the sampled fine-scale permeability field, Bottom left: Posterior density of l, Bottom Right: Posterior density of σ^2.

model resembles the true permeability field. For our example, we consider only the top layer of the five layers in the dataset and follow the same guidelines. We have used the production history i.e the water-cut data, the permeability data on a 5×5 coarse grid and the true fine-scale permeability data only on the well locations to infer about the fine-scale permeability field. The permeability measurements are expressed in the unit of mD where $1mD = 10^{-3}$ Darcy $= 10^{-12}m^2$. We use log transformation of the permeability data and logit transformation of the fractional flow data in our model. The spatial locations of the field were given to the researchers in a transformed Cartesian co-ordinate system with each grid of 180×180 square unit starting from the origin, i.e co-ordinate of the top-left grid block is $(0, 0)$ and that of the bottom-right grid block is $(3420, 5040)$. For simplification we make another transformation on the co-ordinate system to a $(0, 1)$ scale. So in the transformed spatial domain the co-ordinate of the bottom right grid block is $(0.6786, 1)$ and each grid block is of size 0.0357×0.0357 square unit. We use a squared exponential structure for the prior distribution of the fine-scale log-permeability field while doing the Karhunen Loeve transform. We assume a proper prior for the correlation length which is uniform on a truncated space. We draw 200000 samples from the posterior distribution using two-stage reversible jump MCMC method. After 20000 burn in period we retain every 10th sample from the posterior. We can see from the Figure 9 that the posterior median of the fine-scale permeability is very close to the true punq permeability field. The mode of the number of co-efficient to be retained in KLE expansion is found to be 24. The posterior mode of l is nearly 0.25 and that of σ^2 is nearly 9 (see Figures 10 and 11). To visualize the uncertainties in the prediction we plot the first and third quartile of the sampled fine-scale permeability field from the posterior in Figure 12. Next we consider the model assuming no coarse-scale data available. From Figure 13 we can see that the posterior median is not very close to the true punq model. Hence,

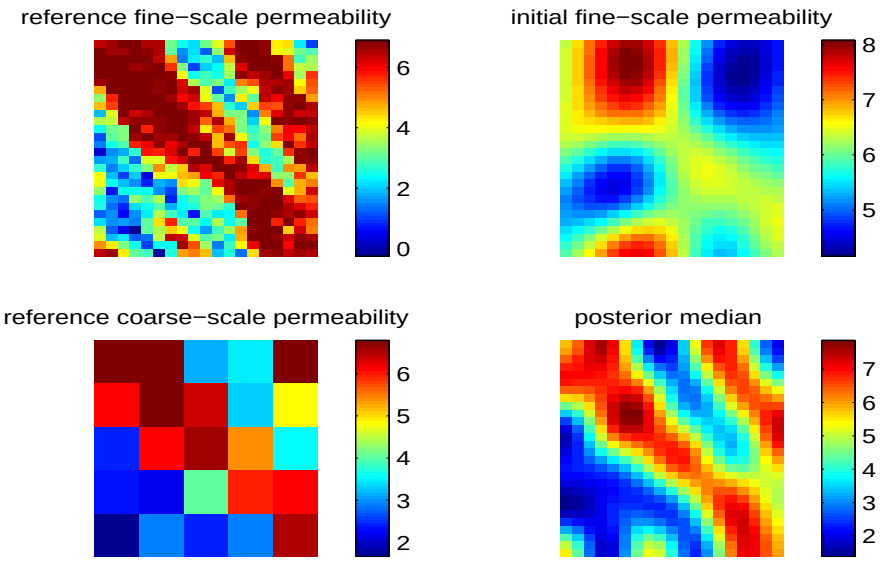

Fig. 9. Log permeability field using the two stage reversible jump MCMC for the PUNQ-S3 model. Top left: The true fine-scale log permeability field, Top right: Initial fine-scale log permeability field, Bottom left: The observed coarse-scale permeability field, Bottom Right: The median of the sampled fine-scale permeability field.

we can conclude that integrating coarse-scale data in the model helps us to predict the uncertainties in the reservoir more efficiently. The sum of squared error between the true fine-scale permeability and the posterior median when we use the available coarse-scale data is 542.2783. In contrast, when we only use the fine-scale data at a few well locations but no coarse-scale data the sum of squared error becomes 1192.4.

II.6. Conclusions

We have developed a Bayesian multiscale hierarchical model for large scale inverse problems. Posterior distribution has been used for uncertainty quantification. Two stage MCMC technique has been exploited for computational efficiency. Alternatively,

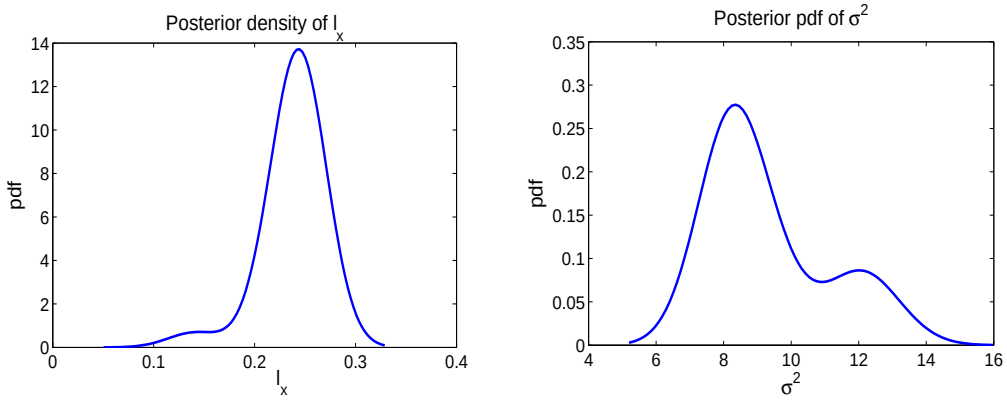

Fig. 10. Posterior density of l and σ^2 for the PUNQ-S3 model. Left: Posterior density of l, Right: Posterior density of σ^2.

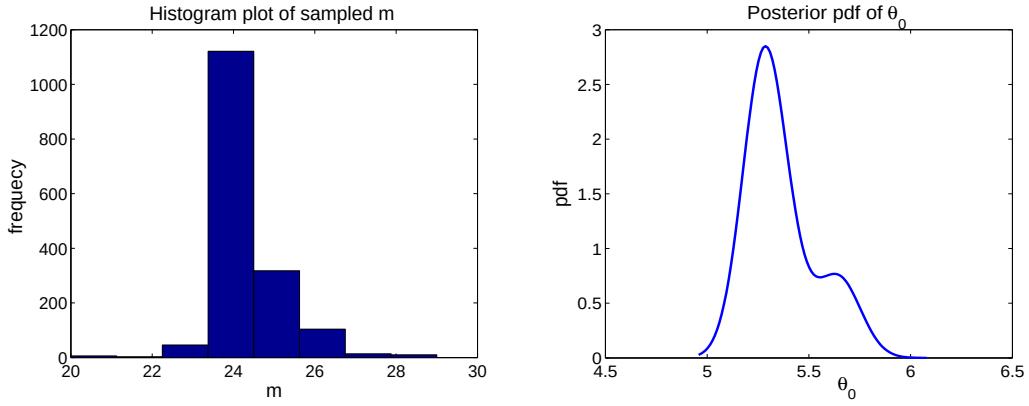

Fig. 11. Posterior distributions of m and θ_0 for the PUNQ-S3 model. Left: Histogram of the posterior distribution of m, Right: Posterior density of θ_0.

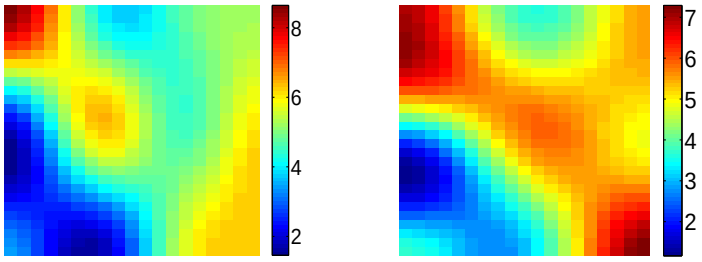

Fig. 12. Quartiles of the sampled log permeability field for the PUNQ-S3 model. Left: The first quartile of the sampled posterior fine-scale log permeability field, Right: The third quartile of the sampled posterior fine-scale log permeability field.

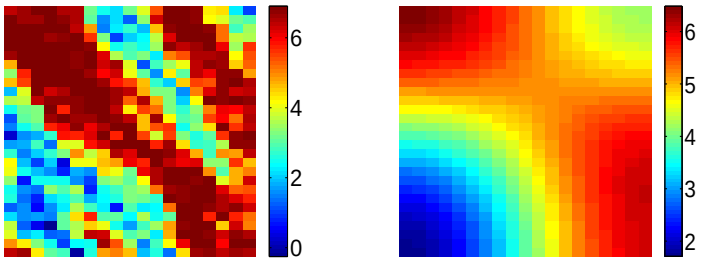

Fig. 13. Log permeability for the PUNQ-S3 model assuming no coarse scale data avaialable. Left: The true fine-scale log permeability field, Right: The median of the sampled fine-scale permeability field.

statistical interpolation techniques like emulator (see Oakley and O'Hagan (2004), Kennedy and O'Hagan (2001), Williams *et al.* (2006) and Higdon *et al.* (2004)) can be used in this problem. Development of multi scale emulator for this problem will be our challenging future project. Furthermore, use of Metropolis-adjusted Langevin algorithm (e.g. Roberts and Rosenthal (2001), Dostert *et al.* (2006)) may be useful to develop more computationally efficient MCMC method.

CHAPTER III

INVERSE PROBLEMS IN SPATIAL FIELDS WITH CHANNELIZED STRUCTURE*

In many geologic environments, the distribution of subsurface properties is primarily controlled by the location and distribution of distinct geologic facies with sharp contrasts in properties across facies boundaries (see Weber (1990)). For example, in a fluvial setting, high permeability channel sands are often embedded in a nearly impermeable background causing the dominant fluid movement to be restricted within these channels. Under such conditions, the orientation of the channels and channel geometry determine the flow behavior in the subsurface rather than the detailed variations in properties within the channels. Traditional geostatistical techniques for subsurface characterization have typically relied on variograms that are unable to reproduce the channel geometry and the facies architecture (see Haldorsen and Damsleth (1990), Koltermann and Gorelick (1996) and Dubrule (1998)). Various other approaches have been applied for modeling facies, e.g., discrete Boolean or object-based models (see Egeland *et al.* (1993)). The success of these object-based models is heavily dependent on the parameters to specify the object size, shapes, proportion and orientation.

Several authors have used the adjustment of paleochannel parameters as a mechanism to match the production data and update the facies models. This approach allows us to take advantage of the gradient based inverse methods but is limited with respect to channel shapes and geometry. For example, Landa and Horne (1997) used

---

*Part of the data reported in this chapter is reprinted with permission from "Bayesian Uncertainty Quantification for Flows in Heterogeneous Porous Media using Reversible Jump Markov Chain Monte Carlo Methods" by A. Mondal, Y. Efendiev, B. Mallick and A. Datta-Gupta, 2010, *Advances in Water Resources*, Volume 33, Issue 3, Pages 241-256, Copyright[2010] by Elsevier

trigonometric functions to model the channel boundaries. The channel boundaries were moved to match the dynamic response but were always kept parallel. This was generalized by Bi *et al.* (2000) to accommodate more flexible channel geometry. The channel shapes and orientations were specified using principal direction, horizontal and vertical sinuosity of the channel and the width and aspect ratio of the channel. However, the use of geologic objects restricted the ability to generate multiple facies architecture. The introduction of truncated plugaussian models allowed for considerable flexibility in terms of facies textures and shapes, for example see Galli *et al.* (1994). The approach requires specification of at least two covariance models and truncation thresholds but allows for multiple facies and a variety of facies association. The conditioning of these models to dynamic data is again complicated by the discrete representation of the facies that makes the application of gradient-based methods difficult and often inefficient (see Liu and Oliver (2005)). Recently geostatistical models based on multi point statistics have been proposed for reproduction of complex channel architectures. These methods rely on training images that can be difficult to obtain. Also, current multi point methods are well suited for subsurface characterization using static data only and do not allow for efficient integration of dynamic data. A rigorous formalism for uncertainty quantification is largely missing in all of the methods discussed above.

In this chapter, our goal is two-fold: (1) hierarchical modeling of spatial fields with channelized architecture; (2) efficient sampling of the posterior probability distribution with hierarchical priors. In hierarchical modeling, the spatial field is represented by facies boundaries and variogram-based spatial fields within each facies. Typically, the parameters representing facies boundaries are highly uncertain, particularly in the early stages of subsurface characterization (see Caumon *et al.* (2004) and Dubrule (1998)). In a channel type environment, the channel sands may be ob-

served at a few well locations. The observations at the well locations can be used in conjunction with reversible jump Markov chain Monte Carlo methods to construct the parameterization of the facies in a computationally efficient manner. This is one of our objectives in this chapter. There are many plausible channel geometries that will satisfy the channel sand distribution, orientation and well intersections. Thus, the stochastic models for channels will require specification of random variables that govern the channel principal direction, its horizontal and vertical sinuosity, channel width to thickness ratio etc. All these parameters have considerable uncertainty associated with them but will profoundly impact fluid flow in the subsurface. In this chapter, the channel boundaries are represented using piecewise linear functions - an approach capable of reproducing a wide variety of channel geometry. The shape of the channel boundaries is updated with dynamic data using reversible jump MCMC where the number of points representing the channel boundaries is assumed to be unknown. In the reversible jump MCMC method the dimension of the parameter space is also taken to be random. Note that in a conventional MCMC method the dimension of the parameter space is fixed. This flexibility allow us to have a birth or a death step at each iteration of the reversible jump MCMC method. In a birth step we add one more point on the channel boundary and thus increase the dimension by one where as in a death step we delete one point on the boundary and thus reduce the dimension by one. We can also have a jump step like in conventional MCMC methods at each iteration of the reversible jump MCMC method. Thus in reversible jump MCMC method we can move the points along the horizontal directions by having a birth or a death step and we can move the points in vertical directions by having a jump step. In standard MCMC method we only have the jump step, so the points on the channel boundaries can only move in vertical directions. So the reversible jump MCMC method is more flexible in spanning all possible channel shapes and

thus allows an efficient search in the uncertainty space.

Within each facies, a variogram based spatial field is used. To represent variogram-based spatial fields, Karhunen-Loéve expansion (see Loève (1977)) is used. Karhunen-Loéve expansion allows significant reduction in the number of parameters for correlated spatial fields. The truncation procedure of the K-L expansion introduce some error in the posterior measure. In this chapter we also estimate a bound for the difference in the expectation of a function with respect to the full and the truncated posterior. Based on this bound, the computation can be simplify by choosing less number of terms in truncation, while the error is in a reasonable range. This is very advantageous in history matching because it allows to perform the search in a smaller parameter space. Because the spatial fields are independent within different channels, the uncertainty space is still quite large. The spatial field is further conditioned at the well locations. The conditioning can be performed within Karhunen-Loéve expansion.

The sampling of the posterior is done using reversible jump Metropolis-Hastings MCMC. Each proposal is screened by running detailed fine-scale models. It turns out that the acceptance rate of this algorithm is very small. To speed-up the algorithm, we employ two-stage MCMC (Efendiev *et al.* (2006, 2007, 2008); Ma *et al.* (2008)) methods, where coarse-scale simulations are used to screen the proposals.

In this chapter, we present the formulation of two-stage reversible jump MCMC which differs from two-stage MCMC proposed earlier because of the associated birth and death processes. The acceptance rate of two-stage MCMC is further improved by using mixed multiscale finite element methods (MsFEM) for preconditioning of reversible jump MCMC methods.

As in the first chapter for the definiteness of the problem our method is applied for reservoir characterization where the unknown quantity are two dimensional channelized permeability field. The channel boundaries are modeled with reversible jump

MCMC where the number of points is assumed to be unknown. Within each channel, the permeability field is characterized by two-point correlation functions. We assume that the values of the permeabilities are known at the wells. We consider simple flow-based upscaling techniques by averaging the permeability field within each channel. This gives very coarse description of the media.

In the numerical results, we use priors for the number of points at the channel interfaces, the locations of these points, death and birth processes, and the permeability fields within each channel. The initial locations of the interfaces are taken un-informative. For example, we take the initial channel boundaries to be flat line segment, while the reference channel has substantial lateral variations. As for the permeability within each channel, we take the initial permeability to be homogeneous permeabilities, while the reference permeabilities are chosen to be heterogeneous. Our numerical results show that the proposed algorithm can adequately predict the boundaries of the channels. Our algorithm produces some small oscillations along the boundaries; however, the main features of the boundaries are correctly predicted. The acceptance rate of reversible jump MCMC is improved by screening the proposal with upscaled models and mixed multiscale finite element methods. In particular, an error model is constructed based on off-line computations of fine- and coarse-scale models to allow for the bias-correction from coarse-scale models.

The chapter is organized as follows. In Subsection III.1 we discuss the parameterization of the channelized spatial field. In Subsection III.2, the Bayesian hierarchical model are described. Subsection III.3 gives an estimate of the bound on the posterior error for the truncation of K-L expansion. The sampling procedure from the posterior are described in Subsection III.4 and Subsection III.5. Finally, in Subsection III.6, we present numerical results.

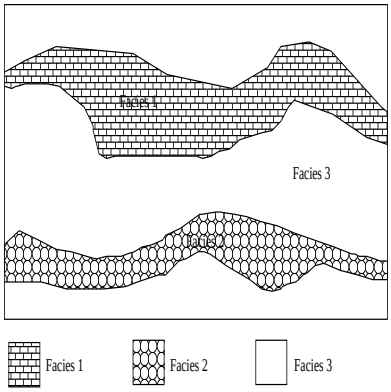

Fig. 14. Illustration of the permeability field with facies.

III.1. Parameterization of the Channelized Spatial Field

We introduce the way we use to parameterize channelized spatial field in this subsection. First, a heterogeneous spatial field is decomposed into several subregions (for example in subsurfaces high and low permeable regions), where each region represent a facies (see Figure 14 for illustration). The spatial field within each faces is assumed to follow a log-Gaussian distribution with a known spatial covariance. This type of hierarchical representation allows us to write of the spatial field as

$$k(x) = \sum_i k_i(x) I_{D_i}(x), \tag{3.1}$$

where I_{D_i} is an indicator function of region D_i (i.e., $I(x) = 1$ if $x \in D_i$ and $I(x) = 0$ otherwise).

In this chapter, we seek the boundaries of the facies using adaptive representation. More precisely, level set functions τ representing the facies boundaries are defined such that $\tau = \tau_i$ for different interfaces. For the update of the facies, the level set equations (e.g., Osher and Fedkiw (2003); Sethian (1999)) will be used. More

precisely, we assume

$$\frac{\partial \tau}{\partial s} + w \cdot \nabla \tau = 0, \tag{3.2}$$

where w is a vector field and s is a pseudo-time. Equation (3.2) is used for the update of the interface. This equation is a linear transport equation where one needs to specify the velocity field w. We take w to be a random divergence-free field with a deterministic flow direction. One can use random forcing instead of random velocity. Because the flow direction is deterministic, Equation (3.2) will be solved using streamline approaches (e.g., Datta-Gupta and King (2007)). Streamline approaches reduce (3.2) into ODE along the characteristic of flow directions. In particular, the streamlines are defined by

$$\frac{dl_s}{ds_x} = w,$$

where s_x is the spatial coordinates along the streamlines. In our simulations, vertical streamlines are used. If we denote by τ the points of the interface, then the update of these points will be given by

$$\tau_{n+1}(l_s(x)) - \tau_n(l_s(x)) = R\delta s_x,$$

where R is a random variable. The equation above is a physical interpretation of the instrumental proposal distribution for τ used on subsequent algorithms.

Next, we discuss the parameterization of $k_i(x)$ within each facies. Within each facies the spatial field is assumed to be loggaussian field. Let us denote the loggaussian field as $Y_i(x) = log(k_i(x)$.

If we discretize the domain D_i by a rectangular mesh, the continuous K-L expansion (Loève (1977)) of the spatial field is reduced to finite terms. The discretized

K-L expansion is given by

$$Y_{N_i} = \sum_{j=1}^{N_i} \sqrt{\lambda_{ij}^{(\theta)}} \theta_{ij} \psi_{ij}(x), \tag{3.3}$$

where N_i is the grid size, i.e $N_i = N_{i_x} \times N_{i_y}$. In our numerical examples, we will use gaussian correlation structure, though the method is not restricted to this particular covariance structure. $R(x, y)$ in this case is defined as

$$R(x, y) = \sigma^2 \exp\Big(-\frac{|x_1 - y_1|^2}{2l_1^2} - \frac{|x_2 - y_2|^2}{2l_2^2}\Big), \tag{3.4}$$

l_1 and l_2 are the correlation lengths in each dimension, and $\sigma^2 = E(Y^2)$ is the variance.

In the numerical experiments, we first generate a reference permeability field using all eigenvectors and compute the corresponding fractional flows. To propose permeability fields from the prior (unconditioned) distribution for the i^{th} facies, we keep M_i terms in the KLE. Suppose the permeability field is known at M_{H_i} distinct points. This condition is imposed by setting

$$\sum_{k=1}^{M_{H_i}} \sqrt{\lambda_k} \theta_k \phi_k(x_j) = \beta_j, \tag{3.5}$$

where β_j $(j = 1, \ldots, M_{H_i})$ are prescribed constants. In this system, we identify M_{H_i} unknowns for which the system will be solved by choosing the rest of $M_i - M_{H_i}$ θ's normally distributed. These unknowns are found by searching all $M_{H_i} \times M_{H_i}$ minors of $M_i \times M_{H_i}$ matrices with the best condition number. Here M_i must be chosen such that M_{H_i} is less than M_i.

III.2. Bayesian Hierarchical Model

Our main objective is to sample the permeability field given fractional flow measurements. We also incorporate the information that the spatial field is known at some

spatial locations corresponding to wells. The output is an integrated response and the map from spatial field to the output is not one-to-one. So there may exist many different spatial field realizations for a given output. The measured output data is denoted as z. For a given spatial field k, we denote the output as $G(k)$. $G(k)$ can be computed by solving the model equation (2.35)-(2.38) on the fine-grid. In case of reservoir models the spatial field k is the permeability field and the output $G(k)$ is the fractional flow or water-cut data. So we take a logit transformation on on the output and write $F(k) = logit(G(k))$. Using Bayes' theorem the posterior distribution can be written as

$$\pi(k) = P(k|z) \propto P(z|k)P(k). \tag{3.6}$$

III.2.1. Modeling the Likelihood $P(z|k)$

The computed $G(k)$ will contain both modeling error and measurement error. Assuming the combined error as a random error ϵ we can write the model as

$$z = logit(G(K)) + \epsilon = F_k + \epsilon, \tag{3.7}$$

where ϵ is distributed as $N(0, \sigma_f^2 I)$. i.e., $P(z|k)$ is assumed to be $N(F_k, \sigma_f^2 I)$.

III.2.2. Modeling the Prior Process $P(k)$

Consider a spatial field $k(x, \omega)$ in D which has s facies $\{D_i\}_{i=1}^s$ and $s-1$ interfaces $\{\tau_i\}_{i=1}^{s-1}$ and boundaries τ_0, τ_s, assuming s is finite, each facies is described by a covariance matrix $R_i(x, y)$ as in Subsection II.1.1. Then the spatial filed $k(x, \omega)$ is a simple function

$$k(x, \omega) = \sum_{i=1}^{s} k_i(x, \omega) I_{D_i}(x).$$

Since the spatial field of each face D_i is given by $k_i(x, \omega) = \exp\{Y_i(x, \omega)\} = \exp\{\sum_{j=1}^{\infty} \sqrt{\lambda_{ij}} \theta_{ij} \psi_{ij}(x)\}$ and each interface is parameterized by τ the spatial field

$k(x,\omega)$ can be also written as

$$k(\theta,\tau) = \sum_{i=1}^{s} \exp(Y_i) I_{D_i(\tau)}(x).$$

Considering discretized space allows us to write Y_i in each D_i as $Y_{N_i} = \sum_{j=1}^{N_i} \sqrt{\lambda_{ij}}\theta_{ij}\psi_{ij}(x)$, $i = 1, \cdots, s$. Notice that λ_i's usually drop to 0 fast, the truncation of K-L expansions, i.e. $Y_{M_i} = \sum_{i=1}^{M_i} \sqrt{\lambda_{ij}^{(\theta)}}\theta_{ij}\psi_{ij}$ can be used to reduce the dimension of the parameter space, which in turn would save a lot of CPU time while sampling from the posterior distribution. We denote $\theta = (\theta_{11}, \cdots, \theta_{1N_1}, \cdots, \theta_{s1}, \cdots, \theta_{sN_s})$ and $\tau = (\tau_1, \cdots, \tau_{s-1})$. Let θ_M denote the truncation of θ then the corresponding representation of the permeability fields are given by:

$$k(\theta_M,\tau) \;=\; \sum_{j=1}^{s} \exp(\sum_{i=1}^{M_j} \theta_{ji}\sqrt{\lambda_{ji}}\psi_{ji}) I_{\{D_j(\tau)\}}. \tag{3.8}$$

So from the parameterization of interfaces with level sets we can say that the permeability field k is completely known given θ and τ. We assume independent gaussian priors for both θ and τ, i.e $\theta \sim N(\theta_0, \sigma_\theta^2)$ and $\tau \sim N(\tau_0, \sigma_\tau^2)$.
So the hierarchical model is given by:

$$\begin{aligned} \pi(k) = P(k|z) &\propto P(z|k)P(k) \\ &= P(z|k)P((\theta,\tau)') \\ &= P(z|k)P(\theta)P(\tau)[\text{since}\,\theta\,\text{and}\,\tau\,\text{are independent}]. \end{aligned} \tag{3.9}$$

III.3. Posterior Error Introduced by Truncation

Let the spatial field representation using all the terms in the discretized K-L expansion is given by: $k(\theta,\tau) = \sum_{j=1}^{s} \exp(\sum_{i=1}^{N_j} \theta_{ji}\sqrt{\lambda_{ji}}\psi_{ji}) I_{\{D_j(\tau)\}}$ and using the truncated version of the K-L expansion is given by $k(\theta_M,\tau) = \sum_{j=1}^{s} \exp(\sum_{i=1}^{M_j} \theta_{ji}\sqrt{\lambda_{ji}}\psi_{ji}) I_{\{D_j(\tau)\}}$
Correspondingly, the two posteriors distribution of permeability field in Bayesian

framework are given by:

$$\begin{aligned}\pi(\theta,\tau) &\propto G(\theta,\tau)\prod_{i=1}^{s}\pi_0(\theta_{i1},\cdots,\theta_{iN_i})\prod_{j=1}^{s-1}\pi_0(\tau_j),\\ \tilde{\pi}(\theta,\tau) &\propto \tilde{G}(\theta_M,\tau)\prod_{i=1}^{s}\pi_0(\theta_{i1},\cdots,\theta_{iN_i})\prod_{j=1}^{s-1}\pi_0(\tau_j),\end{aligned}$$

where $G(\theta,\tau) = \exp(-\frac{||z-F(k(\theta,\tau))||}{\sigma_f^2})$, $\tilde{G}(\theta_M,\tau) = \exp(-\frac{||z-F(k(\theta_M,\tau)||}{\sigma_f^2})$, and z is the observed fractional flow data.

It is obvious that this truncation process affects the matching process. Our goal here is to find an estimation of error introducing by this truncation, which also provide a way to choose M_i for specified requirements.

The following theorem gives a bound for the trucation error for a simple case where the spatial field has only one facies, or the spatial field does not have a channelized structure.

Theorem III.3.1. *Suppose the permeability field $k = \sum_{i=1}^{N}\theta_i\sqrt{\lambda_i}\psi_i$ is a stationary spatial process on a bounded region, with the truncation $\tilde{k} = \sum_{i=1}^{M}\theta_i\sqrt{\lambda_i}\psi_i$, and $f(\theta)$ is square integrable with respect to Gaussian measure, i.e. $\int|f(\theta)|^2\pi(\theta)d\theta<\infty$, then*

$$|E_{\pi(\theta)}[f(\theta)] - E_{\tilde{\pi}(\theta)}[f(\theta)]| \le C\{\sum_{i=M+1}^{N}\lambda_i\}^{\frac{1}{2}}, \tag{3.10}$$

where C is independent of dimension N.

This theorem is proved in the appendix D. Now for the general case, when permeability field has different facies, the following results hold

Corollary III.3.2. *Suppose the discretized K-L expansion of the log permeability field in the i^{th} region is given by $Y_{N_i} = \sum_{j=1}^{N_i}\sqrt{\lambda_{ij}}\theta_{ij}\psi_i(x)$ where all Y_{N_i} is a stationary spatial processes on a bounded region. Also, suppose the truncated K-L expansion of the log permeability field in the i^{th} region is given by $Y_{M_i} = \sum_{i=1}^{M_i}\sqrt{\lambda_{ij}}\theta_{ij}\psi_{ij}$.*

Assume that $f(\theta,\tau)$ is a square integrable function with respect to Gaussian measure, i.e. $\int |f(\theta,\tau)|^2\pi(\theta)\pi(\tau)d\theta d\tau < \infty$, then

$$\left|E_{\pi(\theta,\tau)}[f(\theta,\tau)] - E_{\tilde{\pi}(\theta,\tau)}[f(\theta,\tau)]\right| \leq C_1 \max_{1\leq i\leq s}\left\{\sum_{j=M_i+1}^{N_i}\lambda_{ij}\right\}^{\frac{1}{2}}, \tag{3.11}$$

where C_1 is independent of dimension N_i.

Proof. From Theorem III.3.1 it can be showed that

$$\begin{aligned}\left|E_{\pi(\theta,\tau)}[f(\theta,\tau)] - E_{\tilde{\pi}(\theta,\tau)}[f(\theta,\tau)]\right| &\leq C\int\sum_{i=1}^{s}\left\{\sum_{j=M_i+1}^{N_i}\lambda_{ij}\right\}^{\frac{1}{2}}\pi_0(\tau)d\tau \\ &\leq C\max_{1\leq i\leq s}\left\{\sum_{j=M_i+1}^{N_i}\lambda_{ij}\right\}^{\frac{1}{2}}.\end{aligned}$$

□

III.4. Reversible Jump MCMC

If the dimension of the parameters τ and θ is fixed then we can use standard Metropolis-Hastings algorithm to sample from the posterior distribution . But if we keep the the number of the points which determines the interfaces, τ and their positions in the horizontal direction remains fixed, then the points only jump at fixed positions giving different τ's and hence the resulting interfaces may not capture the actual boundaries of the facies very well. If the number of the points and their positions are allowed to vary along with the jumps in each step then the resulting algorithm would offer more flexibility in terms of channel shapes and smoothness of the channel boundaries. If we vary the number of points that determines the interfaces then the dimension of τ will also change in each step. This jumping between different dimensions in the parameter space can be achieved through reversible jump Markov Chain Monte Carlo methods as proposed in Green (1995). In such case using the hierarchical Bayes' model, the

posterior can be written as

$$\begin{aligned} P(k|z) &\propto P(z|k)P(\theta)P(\tau) \\ &= P(z|k)P(\theta)P(\tau|x^{loc},m)P(x^{loc}|m)P(m). \end{aligned} \tag{3.12}$$

Here m is the number of points considered and x^{loc} denotes the locations of those points in the horizontal direction that determine the interfaces. We use the reversible jump process as a birth and death process. At each step either we add a new point or delete a point or consider only jumps at fixed positions. The dimension of τ may vary in each step but the dimension of θ is always the same.

The algorithm is as follows.

Algorithm: Reversible Jump MCMC as Birth and Death Process, (Green (1995))

Suppose at the nth step we are at τ_n, θ_n, x_n^{loc}, m_n and permeability field k_n. Here x_n^{loc} is a m_n dimensional vector $(x_{n,1}^{loc}, x_{n,2}^{loc} \dots x_{n,m_n}^{loc})'$. We have three options: add a point with probability $p_{m_n}^{add}$; delete a point with probability $p_{m_n}^{del}$; or just propose a new θ and τ and thus have a jump step with the locations remained fixed with probability $p_{m_n}^{j}$, where $p_{m_n}^{add} + p_{m_n}^{del} + p_{m_n}^{j} = 1, \ \forall \ m_n$.

- **Birth Step**. Here we add one point from the remaining points and the proposed m is $m_n + 1$ with probability $p(m|m_n) = p_{m_n}^{add}$. So in this case, $d_{m_n m} = 1$ and $d_{m m_n} = 0$. We generate u from $q(u|x_n^{loc})$ and the proposed locations are given by $x^{loc} = (x_1^{loc}, x_2^{loc} \dots x_{m_n+1}^{loc})' = g_{1_{m_n m}}(x_n^{loc}, u) = g_{m_n m}(x_n^{loc}, u)$, where $g_{1_{m_n m}}(x_n^{loc}, u)$ is a deterministic function. The proposed θ is generated from $q_\theta(\theta|\theta_n)$. The acceptance probability is given by

$$\tau^{b}_{m_n m_n+1}(x^{loc}_n, x^{loc}) = \min\Big\{1,\ \underbrace{\frac{P(z|k)}{P(z|k_n)}}_{\text{likelihood ratio}} \times \underbrace{\frac{q_\theta(\theta_n|\theta)p^{del}_{m_n+1}}{q_\theta(\theta|\theta_n)p^{add}_{m_n}q_{m_n m_n+1}(u|x^{loc}_n)}}_{\text{proposal ratio}} \times \underbrace{\frac{P(\theta)P(\tau|x^{loc})P(x^{loc}|m_n+1)P(m_n+1)}{P(\theta_n)P(\tau_n|x^{loc}_n)P(x^{loc}_n|m_n)P(m_n)}}_{\text{prior ratio}} \times \underbrace{\Big|\frac{\partial g_{m_n m}(x^{loc}_n, u)}{\partial x^{loc}_n \partial u}\Big|}_{\text{Jacobian}}\Big\}. \tag{3.13}$$

- **Death Step**. Here we delete one point from the existing points and the proposed m is $m_n - 1$ with probability $p(m|m_n) = p^{del}_{m_n}$. So here $d_{m_n m} = 0$ and $d_{m m_n} = 1$. The proposed locations are given by the function $(x^{loc}, u^{'}) = (x^{loc}_1, x^{loc}_2 \dots x^{loc}_{m_n-1}, u^{'}) = g_{m_n m}(x^{loc}_n) = (g_{1_{m_n m}}(x^{loc}_n), g_{2_{m_n m}}(x^{loc}_n)$ and $x^{loc}_n = g^{-1}_{m_n m}(x^{loc}, u^{'}) = g_{1_{m_n m}}(x^{loc}, u^{'})$. The proposed θ is generated from $q_\theta(\theta|\theta_n)$. The acceptance probability is given by

$$\tau^{d}_{m_n m_n-1}(x^{loc}_n, x^{loc}) = \min\Big\{1,\ \underbrace{\frac{P(z|k)}{P(z|k_n)}}_{\text{likelihood ratio}} \times \underbrace{\frac{q_\theta(\theta_n|\theta)p^{add}_{m_n-1}q_{m_n-1 m_n}(u^{'}|x^{loc})}{q_\theta(\theta|\theta_n)p^{del}_{m_n}}}_{\text{proposal ratio}} \times \underbrace{\frac{P(\theta)P(\tau|x^{loc})P(x^{loc}|m_n-1)P(m_n-1)}{P(\theta_n)P(\tau_n|x^{loc}_n)P(m_n)}}_{\text{prior ratio}} \times \underbrace{\Big|\frac{\partial g_{m_n m}(x^{loc}_n)}{\partial x^{loc}_n}\Big|}_{\text{Jacobian}}\Big\}. \tag{3.14}$$

- **Jumps at Fixed Location**. Here the number of points and their locations in horizontal direction are fixed, so the algorithm is same as Metropolis Hastings Algorithm as described before. The acceptance probability is given by

$$\tau(k_n, k) = \min\Big\{1,\ \underbrace{\frac{P(z|k)}{P(z|k)}}_{likelihood\,ratio} \times \underbrace{\frac{P(\tau)P(\theta)}{P(\tau_n)P(\theta_n)}}_{prior\,ratio} \times \underbrace{\frac{q_\tau(\tau_n|\tau)q_\theta(\theta_n|\theta)}{q_\tau(\tau|\tau_n)q_\theta(\theta|\theta_n)}}_{proposal\,ratio}\Big\}. \tag{3.15}$$

Figure 15 illustrates how the interfaces changes in a birth, death and the jump step of the reversible jump MCMC procedure.

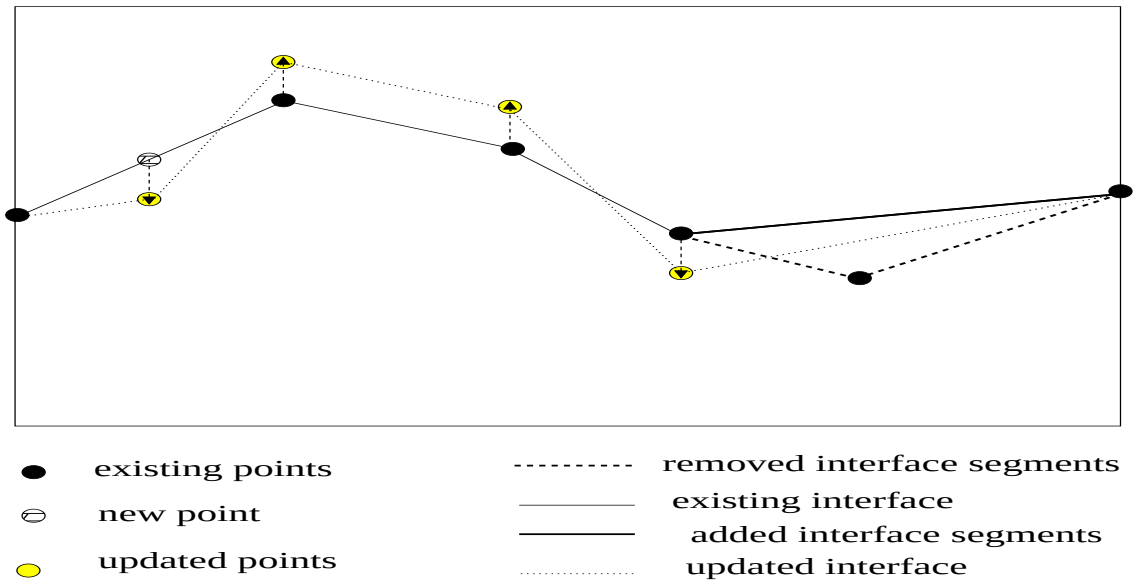

Fig. 15. An illustration of the birth, death and jump process in reversible jump MCMC on an interface.

III.4.1. An Example

As an example, suppose we have a $N \times N$ fine-grid permeability field and let the prior distribution for θ be $N(\theta_o, \sigma_\theta^2)$ and that of τ to be $N(\tau_o, \sigma_\tau^2)$. The prior distribution of m is taken to be discrete uniform distribution, i.e., $P(m) = \frac{1}{(m_{max} - m_{min} + 1)}$, $m = m_{min}\,(1)\,m_{max}$. Given m, the prior distribution of the locations x^{loc} is given by $P(x^{loc}|m) = \frac{1}{(N-2)C_{(m-2)}}$, i.e., $(x^{loc}_{(2)}, x^{loc}_{(3)}, \ldots x^{loc}_{(m-1)})'$ are distributed as order statistics of a sample of size $(m-2)$ drawn without replacement from a population of size $N-2$. Note that, here we have $N-2$ instead of N because the two ends of each of the interfaces are fixed and known and are assumed to be the channel intersections at the wells. Given m and x^{loc}, τ is a vector denoting the heights of the points of the interfaces at given m locations. Also given m, the points for the different interfaces are assumed to be independent. At every step, we join the m points in each interfaces by linear interpolation which completely defines the boundaries of the facies. Then

the reversible jump algorithm will be as follows.

- **Birth Step**. We randomly choose one of the fine-grid interval in $(x^{loc}_{n,(1)}, x^{loc}_{n,(2)}, \ldots x^{loc}_{n,(m_n)})$. Without loss of generality, let the interval be $[x^{loc}_{n,(1)}, x^{uploc}_{n,(2)}]$, Suppose there are C known locations in this interval and the ordered locations are $lx_1, lx_2, \ldots lx_C$. We assign probability $p(i)$ to each of the length $l_i = lx_i - x^{loc}_{n,(2)},\ i = 1\,(1)\,C$. Then we generate one of the lengths from the previously defined probability distribution and add it to $x^{loc}_{n,2}$ to get the newly added location. So the proposed location vector is given by $x^{loc} = (x^{loc}_{(1)}, x^{loc}_{(2)}, \ldots x^{loc}_{(m_n)+1})' = (x^{loc}_{n,(1)}, x^{loc}_{n,(2)} + u_t, x^{loc}_{n,(2)}, \ldots x^{loc}_{n,(m_n)})' = g_{m_n m_n+1}(x^{loc}_n, u)$, where u is drawn from the probability distribution of $l(i)$'s. Hence in this case $q_{m_n m_n+1}(u|x^{loc}_n) = p(i)$, if $u = l(i)$.

  Here we consider

$$p(i) = \frac{P_N\Big(\frac{\frac{(l(i)+l(i+1))}{2} - \frac{(x^{loc}_{n,(1)} - x^{loc}_{n,(2)})}{2}}{\sigma_l}\Big) - P_N\Big(\frac{\frac{(l_i+l_{i-1})}{2} - \frac{(x^{loc}_{n,(1)} - x^{loc}_{n,(2)})}{2}}{\sigma_l}\Big)}{P_N\Big(\frac{\frac{(l_C+l_{C+1})}{2} - \frac{(x^{loc}_{n,(1)} - x^{loc}_{n,(2)})}{2}}{\sigma_l}\Big) - P_N\Big(\frac{\frac{(l(1)+l(0))}{2} - \frac{(x^{loc}_{n,(1)} - x^{loc}_{n,(2)})}{2}}{\sigma_l}\Big)}, \tag{3.16}$$

  $\forall\ i = 1\ (1)\ C$, with $l_0 = x^{loc}_{n,(1)} - x^{loc}_{n,(2)}$ and $l_{C+1} = 0$, where $P_N(\cdot)$ is the standard normal cdf function. We generate u from this probability distribution, which can be considered as a discretized version of normal distribution and which guarantees that the added new point lies in the middle of the two locations with very high probability. We propose $\theta = \theta_n + h_\theta u_\theta$, where u_θ is generated from a $N(0, I)$ distribution. Then the acceptance probability is given by

$$\tau^b_{m_n m_n+1}(x^{loc}_n, x^{loc}) = \min\Big\{1, \text{likelihood ratio}\times\text{prior ratio}\times\text{proposal ratio}\times\text{Jacobian}\Big\}, \tag{3.17}$$

where

$$\text{likelihood ratio} = \frac{\exp\left(\frac{-\|z-F_k\|^2}{2\sigma_f^2}\right)}{\exp\left(\frac{-\|z-F_{k_n}\|^2}{2\sigma_f^2}\right)}, \tag{3.18}$$

$$\text{prior ratio} = \frac{\exp\left(\frac{-\|\theta^*-\theta_o\|^2}{2\sigma_\theta^2}\right)}{\exp\left(\frac{-\|\theta_n-\theta_o\|^2}{2\sigma_\theta^2}\right)} \times \frac{\exp\left(\frac{-\|\tau^*-\tau_o\|^2}{2\sigma_\tau^2}\right)}{\exp\left(\frac{-\|\tau_n-\tau_o\|^2}{2\sigma_\tau^2}\right)} \times \text{prior-multiplier}, \tag{3.19}$$

$$\text{proposal ratio} = \frac{p_{m_n+1}^{del}}{p_{m_n}^{add}} \times \text{proposal-multiplier}, \tag{3.20}$$

$$\text{Jacobian} = 1. \tag{3.21}$$

The prior-multiplier and the proposal-multiplier depend on the number of interfaces considered. If there are $s-1$ interfacies and we assume the locations for different interfacies are independent with each other then, prior-multiplier= $\left(\frac{(m_n-1)}{\sqrt{2\pi}\sigma_\tau(N-m_n)}\right)^{\tilde{I}}$, proposal-multiplier= $\frac{1}{q(u_1)q(u_2)\cdots q(u_{s-1})}$.

- **Death Step**. We randomly choose one of the points in $\{x_{n,(2)}^{loc}, x_{n,(3)}^{loc}, \ldots x_{n,(m_n-1)}^{loc}\}$. Without loss of generality, let the point be $x_{n,(2)}^{loc}$. Suppose there are C_1+C_2+1 known locations within the interval $[x_{n,(1)}^{loc}, x_{n,(3)}^{loc}]$ and let the locations are denoted by $lx_{11}, lx_{12}, \ldots lx_{1C_1}, x_{n,(2)}^{loc}, lx_{21}, lx_{22}, \ldots lx_{2C_2}$. We delete the point $x_{n,(2)}^{loc}$ and hence propose $x^{loc} = (x_{(1)}^{loc}, x_{(3)}^{loc}, \ldots x_{(m_n)}^{loc})'$. Here we take $u' = (x_{(2)}^{loc} - x_{(3)}^{loc})$ and hence in this case
$g_{m_n m_n-1}(x_n^{loc}) = (g_{1_{m_n m_n-1}}(x_n^{loc}), g_{2_{m_n m_n-1}}(x_n^{loc})) = (x_{n,(1)}^{loc}, x_{n,(3)}^{loc}, \ldots x_{n,(m_n)}^{loc}, (x_{n,(2)}^{loc} - x_{n,(3)}^{loc}))'$, and

$$q_{m_n-1 m_n}(u'|x^{loc}) = \frac{P_N\left(\frac{\frac{lx_{1C_1}+x_{n,(2)}^{loc}}{2} - \frac{(x_{n,(1)}^{loc}+x_{n,(3)}^{loc})}{2}}{\sigma_l}\right) - P_N\left(\frac{\frac{x_{n,(2)}^{loc}+lx_{21}}{2} - \frac{(x_{(1)}^{loc}+x_{(3)}^{loc})}{2}}{\sigma_l}\right)}{P_N\left(\frac{\frac{lx_{2C_2}}{2} - \frac{(x_{(1)}^{loc}}{2}}{\sigma_l}\right) - P_N\left(\frac{\frac{lx_{11}}{2} - \frac{x_{(3)}^{loc})}{2}}{\sigma_l}\right)}, \tag{3.22}$$

where $P_N(\cdot)$ is the standard normal cdf function. We propose $\theta = \theta_n + h_\theta u_\theta$, where u_θ is generated from a $N(0, I)$ distribution. The acceptance probability is given by

$$\tau^d_{m_n m_n - 1}(x^{loc}_n, x^{loc}) = \min\Big\{1, \text{likelihood ratio}\times\text{prior ratio}\times\text{proposal ratio}\times\text{Jacobian}\Big\}, \tag{3.23}$$

where

$$\text{likelihood ratio} = \frac{\exp\left(\frac{-\|z-F_k\|^2}{2\sigma_f^2}\right)}{\exp\left(\frac{-\|z-F_{k_n}\|^2}{2\sigma_f^2}\right)}, \tag{3.24}$$

$$\text{prior ratio} = \frac{\exp\left(\frac{-\|\theta-\theta_o\|^2}{2\sigma_\theta^2}\right)}{\exp\left(\frac{-\|\theta_n-\theta_o\|^2}{2\sigma_\theta^2}\right)} \times \frac{\exp\left(\frac{-\|\tau-\tau_o\|^2}{2\sigma_\tau^2}\right)}{\exp\left(\frac{-\|\tau_n-\tau_o\|^2}{2\sigma_\tau^2}\right)} \times \text{prior-multiplier}, \tag{3.25}$$

$$\text{proposal ratio} = \frac{p^{add}_{m_n-1}}{p^{del}_{m_n}} \times \text{proposal-multiplier}, \tag{3.26}$$

$$\text{Jacobian} = 1. \tag{3.27}$$

The prior-multiplier and the proposal-multiplier depend on the number of interfaces considered. If there are $s-1$ interfaces and we assume the locations for different interfaces are independent from each other then, prior-multiplier= $\left(\frac{\sqrt{2\pi}\sigma_\tau(N-m_n+1)}{m_n-2}\right)^{s-1}$ and
proposal-multiplier= $q(u_1)q(u_2)\cdots q(u_{s-1})$.

- **Jumps at fixed locations**. Here the number of points and their locations are fixed so we only have jumps at those fixed locations. We propose $\tau = \tau_n + h_\tau u_\tau$, where u_τ is generated from a $N(0, I)$ distribution. Similarly propose $\theta = \theta_n + h_\theta u_\theta$, where u_θ is also generated from a $N(0, I)$ distribution. This step is same as a simple random walk Metropolis–Hastings step. The acceptance

probability is given by

$$\tau(k_n, k) = \min\Big\{1, \text{likelihood ratio} \times \text{prior ratio} \times \text{proposal ratio}\Big\}, \quad (3.28)$$

where

$$\text{likelihood ratio} = \frac{\exp\left(\frac{-\|z-F_k\|^2}{2\sigma_f^2}\right)}{\exp\left(\frac{-\|z-F_{k_n}\|^2}{2\sigma_f^2}\right)}, \quad (3.29)$$

$$\text{prior ratio} = \frac{\exp\left(\frac{-\|\theta-\theta_o\|^2}{2\sigma_\theta^2} + \frac{-\|\tau-\tau_o\|^2}{2\sigma_\tau^2}\right)}{\exp\left(\frac{-\|\theta_n-\theta_o\|^2}{2\sigma_\theta^2} + \frac{-\|\tau_n-\tau_o\|^2}{2\sigma_\tau^2}\right)}, \quad (3.30)$$

$$\text{proposal ratio} = 1. \quad (3.31)$$

Here we take $p_{m_n}^{add} = p_{m_n}^{del} = p_{m_n}^{j} = \frac{1}{3},\ \forall\ m_n = (m_{min}+1)\ (1)\ (m_{max}-1)$, $p_{m_{min}}^{add} = \frac{2}{3},\ p_{m_{min}}^{del} = 0,\ p_{m_{min}}^{j} = \frac{1}{3}, p_{m_{max}}^{add} = 0,\ p_{m_{max}}^{del} = \frac{2}{3}$ and $p_{m_{max}}^{j} = \frac{1}{3}$.

III.5. Two-stage Reversible Jump MCMC

The main disadvantage of the above reversible jump MCMC algorithm is very high computational cost in solving the coupled nonlinear PDE system (2.35)-(2.38) on the fine-grid to compute F_k in the target distribution $\pi(k)$. Typically, in our simulations, reversible jump MCMC method converges to the steady state after thousands of iterations and the acceptance rate is also very low. A large amount of CPU time is spent on simulating the rejected samples, making the direct (full) reversible jump MCMC simulations very expensive.

The direct reversible jump MCMC method can be improved by adapting the proposal distribution $q(k|k_n)$ to the target distribution using a coarse-scale model. This can be achieved by a two-stage reversible jump MCMC method, where we compare the fractional flow curves on the coarse-grid model, first. If the proposal is accepted by the coarse-scale test, then a full fine-scale computation will be conducted and the

proposal will be further tested as in the direct reversible jump MCMC method. Otherwise, the proposal will be rejected by the coarse-scale test and a new proposal will be generated from $q(k|k_n)$. The coarse-scale test filters the unacceptable proposals and avoids the expensive fine-scale tests for those proposals. The filtering process essentially modifies the proposal distribution $q(k|k_n)$ by incorporating the coarse-scale information of the problem. The algorithm for a general two-stage MCMC method with upscaling was introduced in Efendiev *et al.* (2006). Our hierarchical model can also take an advantage of inexpensive upscaled simulations to screen the proposals. Here we propose to extend the algorithm to two-stage reversible jump MCMC method. Let F_k^* be the fractional flow computed by solving the coarse-scale model of (2.35)-(2.38) for the given k. This is done either with upscaling methods or mixed MsFEM. The fine-scale target distribution $\pi(k)$ is approximated on the coarse scale by $\pi^*(k)$. Here we have

$$\pi(k) \propto \exp\Big(-\frac{\|z - F_k\|^2}{\sigma_f^2}\Big) \times P(k), \tag{3.32}$$

$$\pi^*(k) \propto \exp\Big(-\frac{(G(\|z - F_k^*\|))^2}{\sigma_c^2}\Big) \times P(k), \tag{3.33}$$

where the function G is estimated based on offline computations using independent samples from the prior. More precisely using independent samples from the prior distribution, the permeability fields are generated. Then both the coarse-scale and fine-scale simulations are performed and $\|z - F_k\|$ vs $\|z - F_k^*\|$ are plotted. This scatterplot data can be modeled by

$$\|z - F_k\| = G(\|z - F_k^*\|) + w, \tag{3.34}$$

where w is a random component representing the deviations of the true fine-scale error from the predicted error. Using the coarse-scale distribution $\pi^*(k)$ as a filter,

the two-stage reversible jump MCMC can be described as follows.

Algorithm: Two-stage reversible jump MCMC as Birth and Death Process

Suppose at the n^{th} step we are at τ_n,θ_n, x_n^{loc}, m_n and permeability field k_n.

- Step 1. This step is the same as the reversible jump MCMC method described earlier. The only difference is the fractional flow F_k^* is computed by solving the coarse-scale model. At k_n, generate a trial proposal $\tilde{k}$ from distribution $q(\tilde{k}|k_n)$ the same way as in the reversible jump MCMC described earlier.

- Step 2. Take the proposal as

$$k = \begin{cases} \tilde{k} & \text{with probability } \tau_p(k_n, \tilde{k}), \\ k_n & \text{with probability } 1 - \tau_p(k_n, \tilde{k}). \end{cases}$$

  If we are at Birth Step then the acceptance probability is given by

$$\tau_p(k_n, \tilde{k}) = \min\Big\{1, \underbrace{\frac{P^*(z|\tilde{k})}{P^*(z|k_n)}}_{\text{likelihood ratio}} \times \underbrace{\frac{P(\theta)P(\tau|x^{loc})P(x^{loc}|m_n+1)P(m_n+1)}{P(\theta_n)P(\tau_n|x_n^{loc})P(x_n^{loc}|m_n)P(m_n)}}_{\text{prior ratio}} \times \underbrace{\frac{q_\theta(\theta_n|\theta)p_{m_n+1}^{del}}{q_\theta(\theta|\theta_n)p_{m_n}^{add}q_{m_nm_n+1}(u|x_n^{loc})}}_{\text{proposal ratio}} \times \underbrace{\Big|\frac{\partial g_{m_nm}(x_n^{loc},u)}{\partial x_n^{loc}\partial u}\Big|}_{\text{Jacobian}}\Big\}. \tag{3.35}$$

  Note that, $P(\tau)P(\theta)P(x^{loc}|m_n+1)P(m_n+1)$ is the same as the prior probability $P(\tilde{k})$ as defined in (3.33).

If we are at death step then the acceptance probability is given by

$$\tau_p(k_n,\tilde{k}) = \min\Big\{1,\ \underbrace{\frac{P^*(z|\tilde{k})}{P^*(z|k_n)}}_{\text{likelihood ratio}} \times \underbrace{\frac{P(\theta)P(\tau|x^{loc})P(x^{loc}|m_n-1)P(m_n-1)}{P(\theta_n)P(\tau_n|x_n^{loc})P(x_n^{loc}|m_n)P(m_n)}}_{\text{prior ratio}} \times \underbrace{\frac{q_\theta(\theta_n|\theta)p^{add}_{m_n-1}q_{m_n-1m_n}(u^{'}|x^{loc})}{q_\theta(\theta|\theta_n)p^{del}_{m_n}}}_{\text{proposal ratio}} \times \underbrace{\Big|\frac{\partial g_{m_nm}(x_n^{loc})}{\partial x_n^{loc}}\Big|}_{\text{Jacobian}}\Big\}. \tag{3.36}$$

If we are going to have jumps at fixed locations then the acceptance probability is given by

$$\tau_p(k_n,\tilde{k}) = \min\Big\{1,\ \underbrace{\frac{P^*(z|\tilde{k})}{P^*(z|k_n)}}_{\text{likelihood ratio}} \times \underbrace{\frac{P(\tau)P(\theta)}{P(\tau_n)P(\theta_n)}}_{\text{prior ratio}} \times \underbrace{\frac{q_\tau(\tau_n|\tau)q_\theta(\theta_n|\theta)}{q_\tau(\tau|\tau_n)q_\theta(\theta|\theta_n)}}_{\text{proposal ratio}}\Big\}. \tag{3.37}$$

Therefore, the final proposal k is generated from the effective instrumental distribution

$$Q(k|k_n) = \tau_p(k_n,k)q(k|k_n) + \Big(1 - \int \tau_p(k_n,k)q(k|k_n)dk\Big)\delta_{k_n}(k). \tag{3.38}$$

In our chapter, we use a simple relation for modeling coarse- and fine-scale errors. In particular, G is taken to be a linear function with the condition $G(0) = 0$. Then our $\pi^*(k)$ becomes

$$\pi^*(k) \propto \exp\Big(-\frac{\|z - F_k^*\|^2}{\sigma_c^2}\Big) \times P(k), \tag{3.39}$$

i.e., on the coarse-scale $z|k$ is assumed to follow $N(F_k^*, \sigma_c^2 I)$ distribution, i.e.,

$$P^*(z|k) \propto \exp\Big(-\frac{\|z - F_k^*\|^2}{\sigma_c^2}\Big), \tag{3.40}$$

where σ_c is the precision associated with the coarse-scale model. The parameter σ_c plays an important role in improving the acceptance rate of the precondi-

tioned MCMC method. The optimal value of σ_c depends on the correlation between $\|F - F_k\|$ and $\|F - F_k^*\|$, which can be estimated by offline computations.

- Step 3. Accept k as a sample with probability

$$\tau_f(k_n, k) = \min\left(1, \frac{Q(k_n|k)\pi(k)}{Q(k|k_n)\pi(k_n)}\right). \tag{3.41}$$

The acceptance probability (3.41) can be simplified as

$$\tau_f(k_n, k) = \min\left(1, \frac{\pi(k)\pi^*(k_n)}{\pi(k_n)\pi^*(k)}\right). \tag{3.42}$$

Assuming that on the fine-scale $z|k$ follows a $N(F_k, \sigma_f^2 I)$ distribution, i.e.,

$$P(z|k) \propto \exp\Big(-\frac{\|z - F_k\|^2}{\sigma_f^2}\Big), \tag{3.43}$$

the acceptance probability (3.42) becomes

$$\tau_f(k_n, k) = \min\left(1, \frac{\exp\Big(-\frac{\|z-F_k\|^2}{\sigma_f^2}\Big)\exp\Big(-\frac{\|z-F_{k_n}^*\|^2}{\sigma_c^2}\Big)}{\exp\Big(-\frac{\|z-F_{k_n}\|^2}{\sigma_f^2}\Big)\exp\Big(-\frac{\|z-F_k^*\|^2}{\sigma_c^2}\Big)}\right). \tag{3.44}$$

In the above algorithm, if the trial proposal $\tilde{k}$ is rejected by the coarse-scale test (Step 2), k_n will be passed to the fine-scale test as the proposal. Since $\tau_f(k_n, k_n) \equiv 1$, no further (fine-scale) computation is needed. Thus, the expensive fine-scale computations can be avoided for those proposals which are unlikely to be accepted. In comparison, the regular reversible jump MCMC method requires a fine-scale simulation for every proposal k, even though most of the proposals will be rejected at the end. Since the computation of the coarse-scale solution is very cheap, Step 2 of the preconditioned MCMC method can be implemented very fast to decide whether or not to run the fine-scale simulation. The second step of the algorithm serves as a

filter that avoids unnecessary fine-scale runs for the rejected samples. It is possible that the coarse-scale test may reject an individual sample which will otherwise have a (small) probability to be accepted in the fine-scale test.

We can use the same illustrating example as presented in subection III.6 and the numerical results shows how the two-stage reversible jump becomes more efficient in terms of CPU. While using this example in the two-stage algorithm in Step 1 we add a new location or delete a location or consider jumps as given locations in the same way as we did in reversible jump MCMC method. In Step 2, the acceptance probability for the birth step, death step, and jumping step remains the same as in (3.17), (3.23) and (3.28) respectively with F_k, F_{k_n} and σ_f^2 in the *likelihood ratio* replaced by F_k^*, $F_{k_n}^*$ and σ_c^2 respectively.

III.6. Numerical Results

In our first numerical example, we consider a 50×50 fine-scale permeability field on the unit square. We consider the case with only one high conductivity layer. Thus there are two interfaces, one for the upper interface and one for the lower interface. The permeability field is known at 8 locations along $x = 0$ and $x = 1$ boundaries. The ends of the interface are fixed at 0.4 and 0.6. One injection well at $(0, 0.5)$ and one production well at $(1, 0.5)$ are placed. Two-phase flow model with quadratic relative permeabilities $k_{rw} = S^2$ and $k_{ro} = (1-S)^2$ are considered. The log of the permeability field within the channel (middle facies) is assumed to be Gaussian process with mean 3 and covariance function given by (3.4), where $l_1 = 0.3, l_2 = .1$ and $\sigma^2 = .32$. The log of the permeability field outside the high conductivity is assumed to be Gaussian process with mean 0 and the same covariance function, where $l_1 = .2, l_2 = .2$ and $\sigma^2 = .32$. We retain the first 20 terms in the KLE. Initially τ's are taken to be equidistant points on the straight line joining the two ends of the interfaces. We

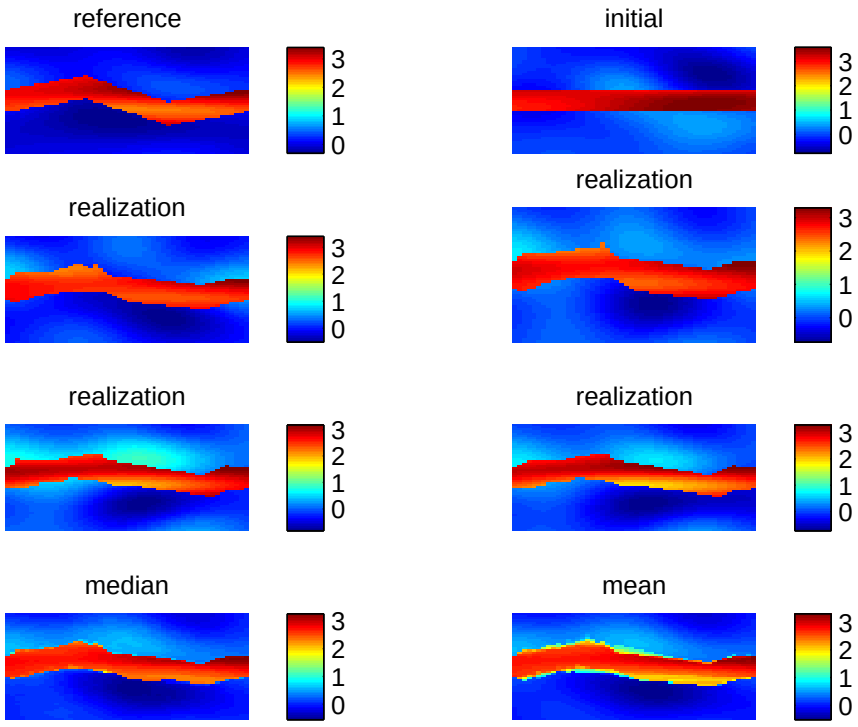

Fig. 16. Log permeability field from full reversible jump MCMC. Top left: The true (reference) log permeability field. Top right: Initial log permeability field. Middle four figures: Accepted realizations of log permeability field. Bottom left: The log of the median of the sampled permeability field. Bottom Right: The log of the mean of the sampled permeability field.

first run the full reversible jump MCMC taking $\sigma_\theta^2 = 0.16$ and $\sigma_\tau^2 = 0.04$. The acceptance rate of the full reversible jump MCMC is very low, approximately 0.002, using $\sigma_f^2 = .004$. The results are shown in Figures 16 and 17.

Next we run several markov chains to sample from the truncated posterior where truncation is done with 10, 15, 20 and 25 terms retained in K-L expansion. The Monte Carlo integration retaining all the terms in the discrete K-L expansion is considered

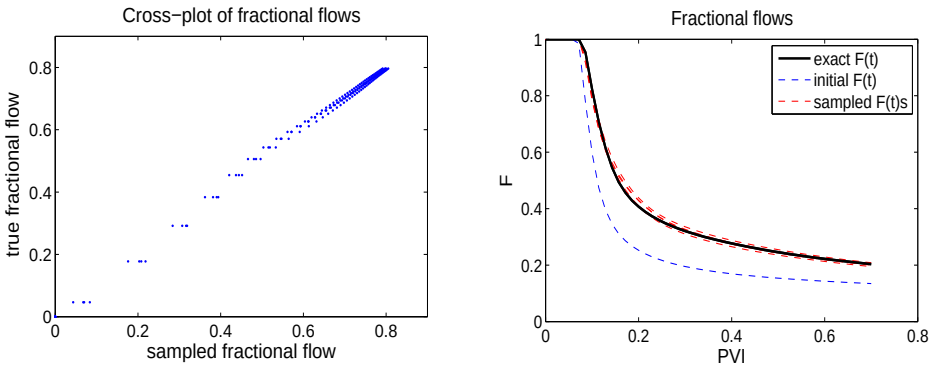

Fig. 17. Fractional flow from full reversible jump MCMC. Left: Cross-plot between the reference fractional flow and sampled fractional flows. Right: Solid black line designates the fine-scale reference fractional flow, the dashed blue line designates the initial fractional flow and the dashed red line designates the fractional flow corresponding to sampled permeability fields.

to be true value of $\int f(\theta)\pi(\theta)d\theta$, and samples with different number of truncated terms are taken to compute $\int f(\theta)\widetilde{\pi}(\theta)d\theta$ as well as comparing with the true one. In Table I we can see that the square error between the true value and the estimated value from the truncated posterior decreases consistently as we increase the number of terms retained in K-L expansion.

Next, we implement two-stage reversible jump MCMC to increase the acceptance rate. Because of mild variations within the facies, one can take volume average of the permeability and avoid more costly single-phase upscaling. We consider two cases. In the first case, the permeability is upscaled via simple volume averaging to three coarse blocks corresponding to facies. In the second example, we divide the domain into 3 equal vertical parts and upscale the permeability within each of 9 blocks. We find such simple and very coarse upscaling works well (i.e., improve the acceptance probability substantially) for the cases where the permeability does not vary too much within facies. The efficiency of these simple upscaling techniques deteriorates as we

Table I. Posterior errors when the K-L expansion is truncated to M terms for two interfaces example.

# of KL terms	$\sqrt{\sum_{i=M+1}^{N} \lambda_i^{(\theta_1)}}$	$\sqrt{\sum_{i=M+1}^{N} \lambda_i^{(\theta_2)}}$	Integration errors
10	0.2959	0.2799	0.0183
15	0.2081	0.1892	0.0153
20	0.1428	0.1258	0.0138
25	0.0991	0.0868	0.0021

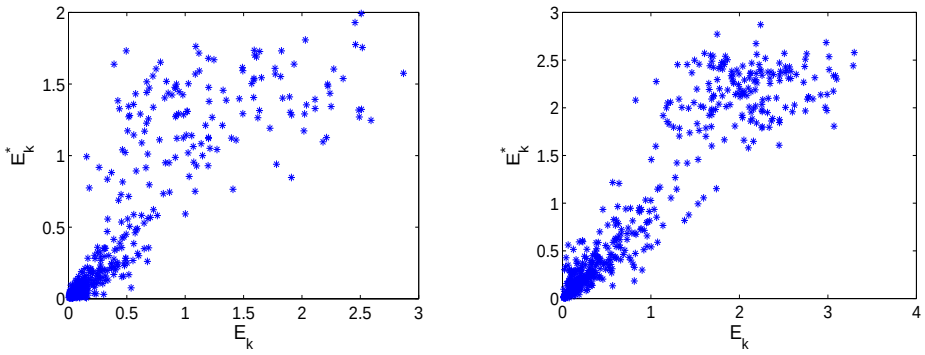

Fig. 18. Cross-plot between $E_k = \|F_{obs} - F_k\|$ and $E_k^* = \|F_{obs} - F_k^*\|$. Left: Cross-plot using three-coarse-block case. Right: Cross-plot using nine-coarse-block case.

increase the variance within the facies. One can improve these methods by taking coarse grid blocks at the kink points of the interface. We have not implemented these coarsening approaches. We suggest the use of mixed MsFEM for the cases with high permeability variations within the facies.

To assess the accuracy of two-stage MCMC, we perform coarse-scale vs. fine-scale simulations for permeability samples from the prior. More precisely, the cross-plot between $E_k = \|F_{obs} - F_k\|$ and $E_k^* = \|F_{obs} - F_k^*\|$ for both three-coarse-block and

nine-coarse-block cases is shown in Figure 18. The correlation coefficient between the two errors are 0.86 and 0.93 for three-coarse-block and nine-coarse-block cases, respectively. This correlation coefficient decreases if the variance of the permeability within facies increases. The high correlation coefficient provides a favorable results for two-stage MCMC. The acceptance rate for the two-stage reversible jump MCMC in the cases of three-coarse-block and nine-coarse-block are 0.33 and 0.47, respectively, using $\sigma_f^2 = 0.004$ and $\sigma_c^2 = 0.01$. In Figure 19, the reference log permeability field, the initial log permeability field, some of the sampled log permeability field, the log mean and median of the sample permeability field for the two stage reversible jump MCMC are shown. We can see that the sampled permeability fields are very close to the reference permeability field. On the left of Figure 20, we depict the cross-plot of fractional flows corresponding to the right figure. In Figure 20 (right plot), we plot the initial fractional flow and the fractional flow corresponding to one of the sampled permeability fields. We observe substantial improvement in fractional flow predictions.

The convergence of two-stage MCMC is plotted in Figure 21. It is clear from this figure, that both two-stage and fine-scale reversible jump MCMC have similar convergence properties, i.e., they reach to the steady state within the same number of iterations. The formal convergence diagnosis can be performed using multiple chains method-based convergence diagnosis (Gelman and Rubin (1992)). In this chapter, our goal is to compare two-stage and direct reversible jump MCMC. We restrict ourselves to only showing errors vs. the number of iterations. We note that the convergence diagnostics has nothing to do with the rate of convergence, which depends on the second largest eigenvalue of the transition matrix of the Markov chain. For the complex chains, the calculation of these eigenvalues is not simple.

In our next numerical example, the same setup is chosen except the variance of

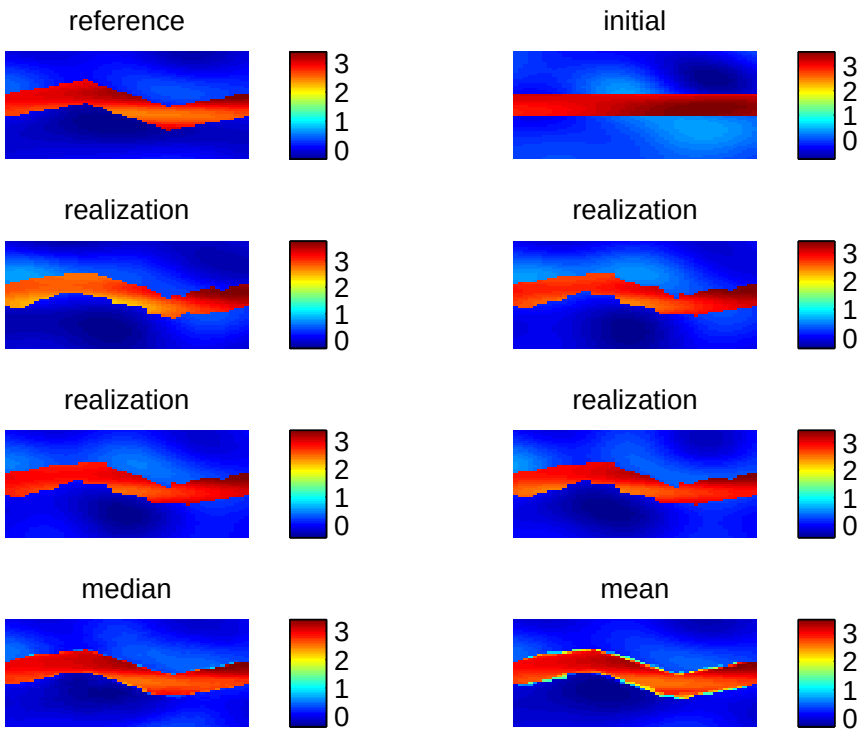

Fig. 19. Log permeability field from two-stage reversible jump MCMC in three–coarse-block case. Top left: The true log permeability field. Top right: Initial log permeability field. Middle four figures: Four accepted realizations of log permeability field. Bottom left: The log of the median of the sampled permeability field. Bottom Right: The log of the mean of the sampled permeability field.

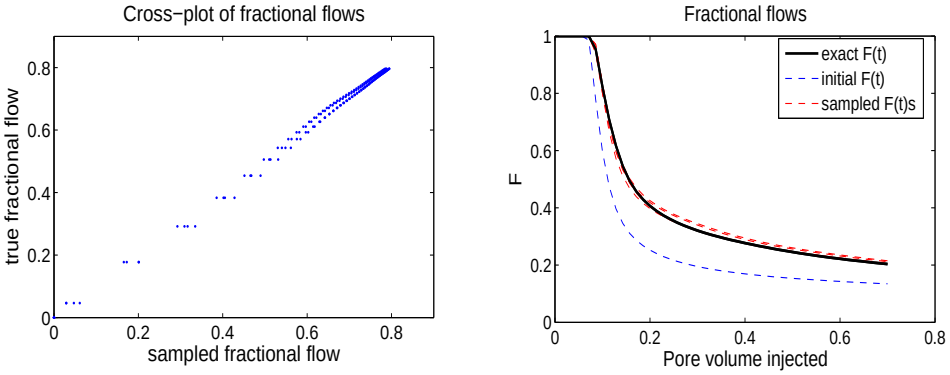

Fig. 20. Cross plot and frcational flow from two-stage reversible jump MCMC in three--coarse-block case. Left: Cross-plot between the reference fractional flow and sampled fractional flows. Right: Solid black line designates the fine-scale reference fractional flow, the dashed blue line designates the initial fractional flow and the dashed red line designate fractional flow corresponding to sampled permeability fields.

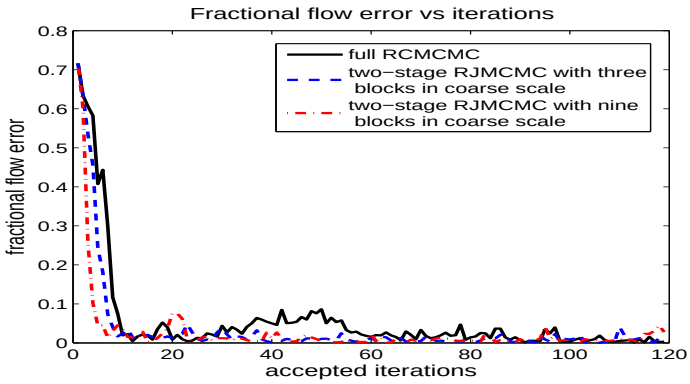

Fig. 21. Fractional flow errors vs. accepted iterations for two-stage and full reversible jump MCMC.

the log permeability field is increased to $\sigma_f^2 = 2$. The full reversible jump MCMC performs as before However, the acceptance rate becomes very low, approximately 0.001. For two-stage reversible jump MCMC, the correlation between $E_k = \|F_{obs} - F_k\|$ and $E_k^* = \|F_{obs} - F_k^*\|$ becomes low, 0.43 and 0.46 for three-coarse-block and nine-coarse-block cases, respectively (see Figure 22). In this case, mixed MsFEM is preferred. The correlation between $E_k = \|F_{obs} - F_k\|$ and $E_k^* = \|F_{obs} - F_k^*\|$ is very high, approximately 0.99, when mixed MsFEM is used (see Figure 22). With mixed MsFEM, the acceptance rate of two-stage reversible jump MCMC increases to 0.31. In Figure 23, we plot the permeability fields obtained using two-stage algorithm. The corresponding fractional flows are plotted in Figure 24. We plot the fractional flow errors vs. iteration number in Figure 25.

In our next set of numerical examples, we consider two high conductivity facies. Thus, there are four interfaces. We assume permeabilities are known in the middle of each facies along $x = 0$ and $x = 1$. As before, the ends of the facies are assumed to be fixed. Two injection wells at $(0, 0.4)$ and $(0, 0.75)$ and two production wells at $(1, 0.4)$ and $(1, 0.75)$ are placed. Two-phase flow model with quadratic relative permeabilities $k_{rw} = S^2$ and $k_{ro} = (1 - S)^2$ are considered. The log of the permeability field inside high conductivity regions is assumed to be Gaussian process with mean 3 and covariance function given by (3.4), where $l_1 = 0.3$, $l_2 = 0.1$ and $\sigma^2 = .32$. The log of the permeability field outside the high conductivity facies is assumed to be Gaussian process with mean 0 and the correlation lengths $l_1 = 0.2$, $l_2 = 0.2$ and $\sigma^2 = 0.32$. The rest of the set up is same as the first example. As before, the full reversible jump MCMC still predicts the interfaces quite accurately. The acceptance probability of reversible jump MCMC is very low, nearly 0.001. The two-stage reversible jump MCMC with five spatial coarse blocks (corresponding to facies) speeds up the process with acceptance rate nearly 0.63 without sacrificing the convergence.

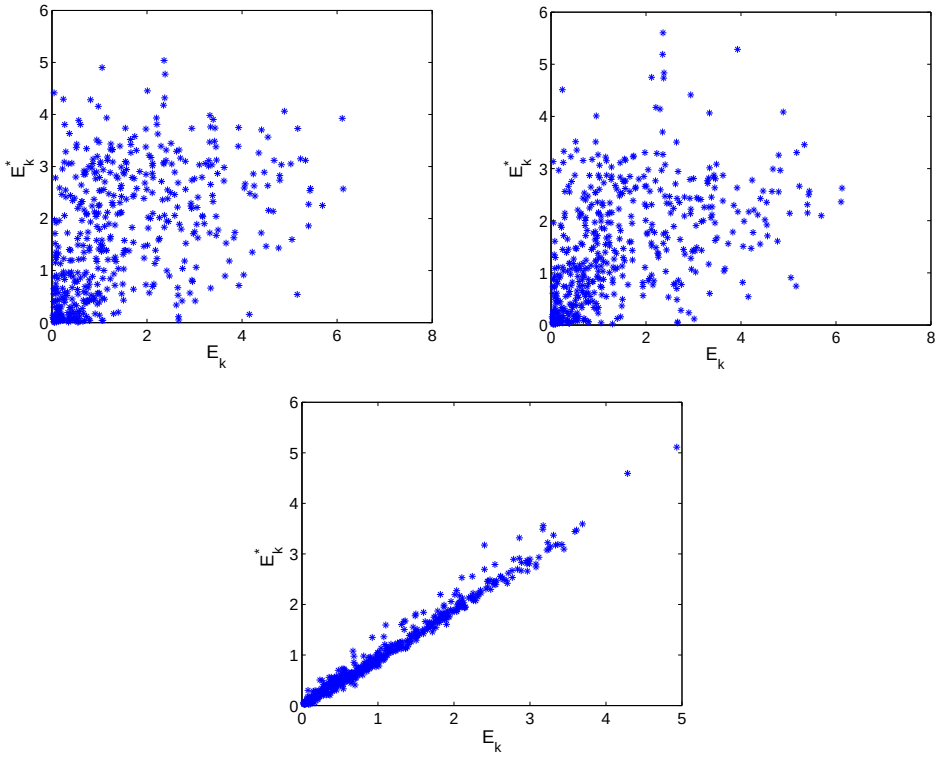

Fig. 22. Cross-plot between $E_k = \|F_{obs} - F_k\|$ and $E_k^* = \|F_{obs} - F_k^*\|$ when the variance of the log permeability field is 2. Left: Cross-plot using three-coarse-block case. Middle: Cross-plot using nine-coarse-block cases. Right: Cross-plot using mixed MsFEM.

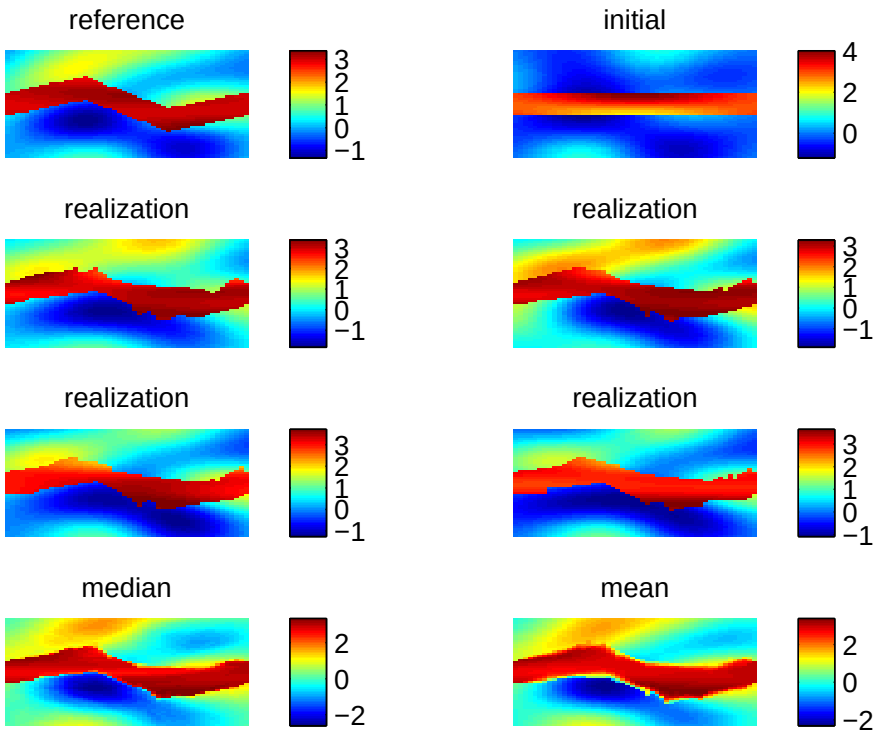

Fig. 23. Log permeability field from two-stage reversible jump MCMC using mixed MsFEM when the variance of the log permeability field is 2. Top left: The true log permeability field. Top right: Initial log permeability field. Middle Four: Four accepted realizations of log permeability field. Bottom left: The log of the median of the sampled permeability field. Bottom Right: The log of the mean of the sampled permeability field.

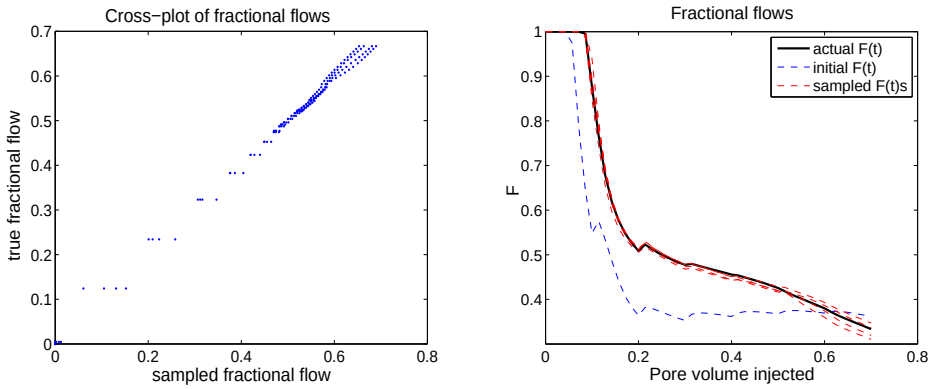

Fig. 24. Corss-plot and frcational flow from two-stage reversible jump MCMC using mixed MsFEM. Left: Cross-plot between the reference fractional flow and sampled fractional flows from two-stage reversible jump MCMC using mixed MsFEM. Right: Solid black line designates the fine-scale reference fractional flow, the dashed blue line designates the initial fractional flow and the dashed red line designate fractional flow corresponding to sampled permeability fields.

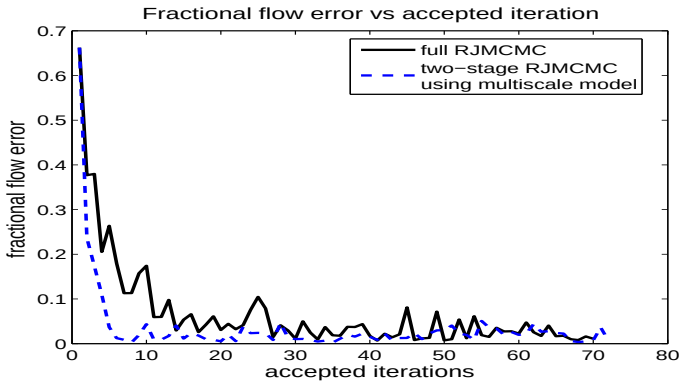

Fig. 25. Fractional flow errors vs. accepted iterations when the variance of the log permeability field is 2.

In Figure 26, we plot the permeability fields obtained using two-stage algorithm. As we see, the prediction is quite accurate in the two-stage reversible jump MCMC as the sampled permeability fields resemble the reference permeability field very closely. The corresponding fractional flows are plotted in Figure 27. Note that there is a substantial improvement in fractional flows when comparing the initial sample and a sample from the posterior. Finally, in Figure 28, we present the fractional flow errors vs. the number of iterations to demonstrate that two-stage reversible jump MCMC has similar convergence as fine-scale reversible jump MCMC.

We again run several markov chains for the four interfaces example to sample from the truncated posterior where truncation is done with 10, 15, 20 and 25 terms retained in K-L expansion. As described before samples with different number of truncated terms are taken to compute $\int f(\theta)\widetilde{\pi}(\theta)d\theta$ as well as comparing with the true one. The square error between the true value and the estimated value from the truncated posterior is shown in Table II. We can see that the truncation error decreases consistently as we increase the number of terms retained in K-L expansion.

Table II. Posterior errors when the K-L expansion is truncated to M terms for four interfaces example.

# of KL terms	$\sqrt{\sum_{i=M+1}^{N}\lambda_i^{(\theta_1)}}$	$\sqrt{\sum_{i=M+1}^{N}\lambda_i^{(\theta_2)}}$	$\sqrt{\sum_{i=M+1}^{N}\lambda_i^{(\theta_3)}}$	Integration errors
10	0.2819	0.2812	0.2722	0.0203
15	0.2011	0.1991	0.1832	0.0181
20	0.1321	0.1241	0.1138	0.0149
25	0.0951	0.0896	0.0823	0.0027

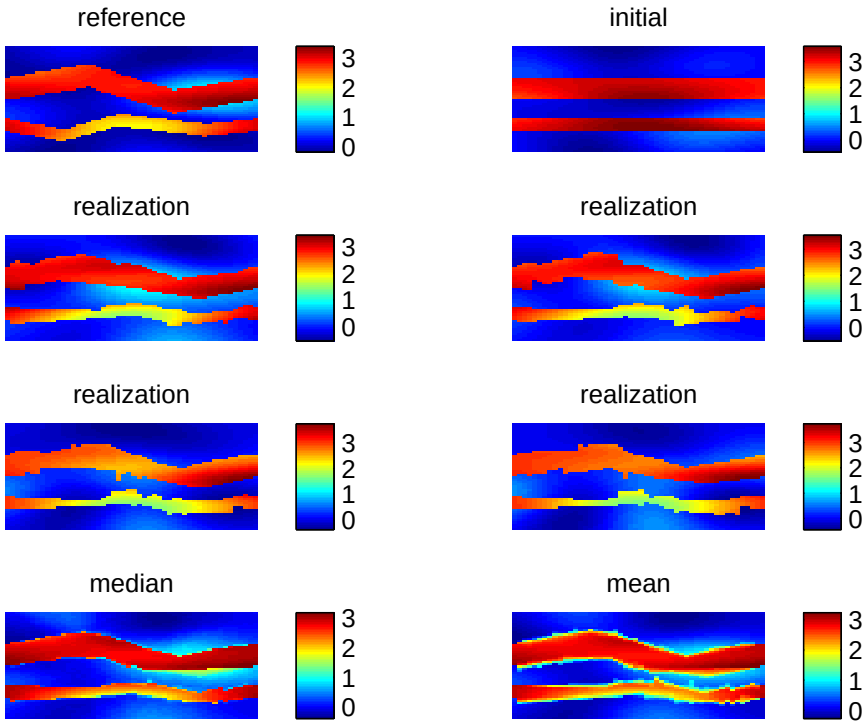

Fig. 26. Log permeability field from two-stage reversible jump MCMC with five coarse blocks. Top left: The true log permeability field. Top right: Initial log permeability field. Middle Four: Four accepted realizations of log permeability field. Bottom left: The log of the median of the sampled permeability field. Bottom Right: The log of the mean of the sampled permeability field.

III.7. Conclusions

In this chapter, we study uncertainty quantification in inverse problems for heterogeneous spatial fiels where the spatial fields have channelized structure. Hierarchical models are used to model the channel boundaries as well as the spatial distribution within the channels that are assumed to be independent. We assume that the channel information at the wells are known; however, no other information is assumed to be

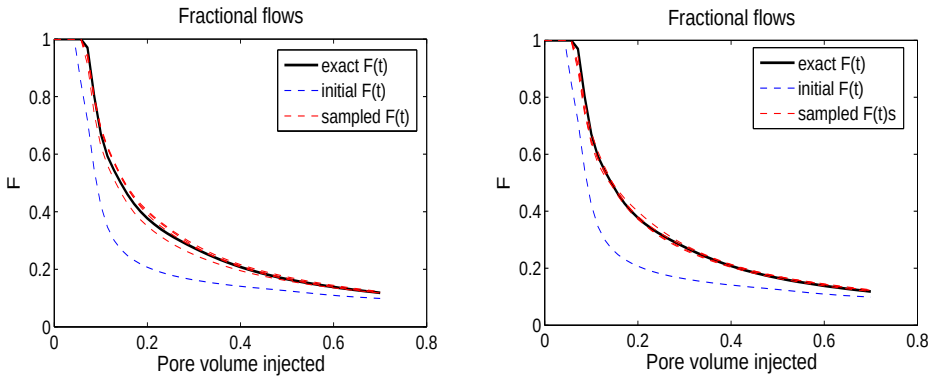

Fig. 27. Comaprison of fractional flow using full reversible jump MCMC vs two-stage reversible jump MCMC in three-coarse-block case. Left: Solid black line designates the fine-scale reference fractional flow, the dashed blue line designates the initial fractional flow and the dashed red line designates fractional flow corresponding to sampled permeability fields from full reversible jump MCMC. Right: Solid black line designates the fine-scale reference fractional flow, the dashed blue line designates the initial fractional flow and the dashed red line designates fractional flow corresponding to sampled permeability fields from two-stage reversible jump MCMC in three-coarse-block case.

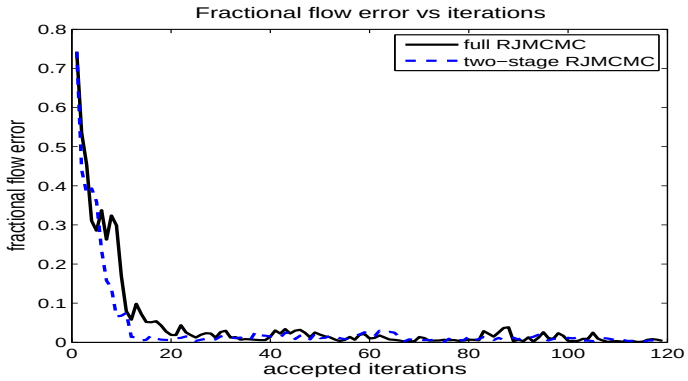

Fig. 28. Fractional flow errors vs. accepted iterations for the example with two channels.

given about the channel shape. The channel boundaries are modeled with variable number of points resulting to changing dimension in the uncertainty space. Reversible jump Markov chain Monte Carlo algorithms are used in such modeling. Within each channel, the spatial field is assumed to have a log-normal distribution. The search with Metropolis-Hastings algorithm results to very low acceptance rate, and consequently, the computations are CPU demanding. To speed-up the computations, we use coarse-scale models to screen the proposals. Our computations show that the proposed algorithms are capable of capturing the channel boundaries and result to accurate predictions of subsurface properties.

In future, Langevin proposals will be used to improve the algorithms. Langevin proposals employ gradient information in making new proposals (Ma *et al.* (2008)). To use the gradient information, we will split the jump process into two parts: (1) adding/deleting new point; (2) perturbing the channel boundaries. Langevin proposal will be computed in the second stage that will provide an easy implementation of two-stage MCMC. Moreover, we will use gradient information based on coarse-scale models for computing the gradients. More precisely in Step 2 of the two-stage reversible jump MCMC algorithm we would choose the proposal generator $q(\tilde{k}|k_n)$ as

$$\tilde{k} = k_n + \frac{\Delta_{(\tau,\theta)}}{2}\nabla \log \pi^*(k_n) + \sqrt{\Delta_{(\tau,\theta)}}\epsilon_n, \tag{3.45}$$

where ϵ_n are independent Gaussian vectors. This will further speed-up the computations.

CHAPTER IV

EMULATORS ON LARGE SCALE SPATIAL INVERSE PROBLEMS

Mathematical models are often used in many areas of science and engineering to describe some physical process over a spatial field. Given the input spatial field and other model input parameters the complex physical process can be modeled by running a numerical simulator, usually called forward operator that uses discretized approximation to a system of non-linear partial differential equations. Often the mathematical models are highly complex so that the computer code can is very expensive in terms of CPU time required for a single run. The spatial inverse problems consists of inferring about the model input parameters given the output observations. Classical statistical approaches to inverse problems have used regularization methods to impose well-posedness, and solve the resulting deterministic problems by optimization and other means Vogel (2002). Here we focus on the Bayesian approach, which contains a natural mechanism for regularization in the form of prior information and which casts the inverse solution as a posterior probability distribution over the model parameters. The posterior distribution of the input model parameters of the system, in most of the cases, is intractable and we often use Markov Chain Monte Carlo method to sample from the posterior. The likelihood term in the posterior contains the forward simulator which need to be computed at each iteration of the MCMC step which is computationally very expensive due to the complexity of the system. So, the Monte Carlo sampling approach rely on large number of execution of the complex forward simulator as thousands of samples from the posterior distribution are needed. Several attempts at accelerating Bayesian inference in inverse problems have been made as an alternative to the direct Markov Chain Monte Carlo technique, for example, Efendiev *et al.* (2006) and Efendiev *et al.* (2007) uses two stage Markov Chain Monte Carlo

technique where the proposals are screened in the first stage using the forward solver in a upscaled coarse grid, which is inexpensive due to small dimensions of the coarse grid. But still we have to run the expensive forward simulator thousands of iterations when the proposals are accepted in the first stage. Another very popular approach to the Bayesian inverse problem is based on polynomial chaos expansion (see Ghamnem and Spanos (1991), Reagan *et al.* (2003), Marzouk *et al.* (2007) and Marzouk and Najm (2009)) where spectral expansion based on Hermite orthogonal polynomials is used to propagate the prior uncertainty through the forward model by a polynomial approximation of the forward simulator over the support of the prior. They have used stochastic Galerkin method to derive an equivalent system of deterministic equations of the original forward model, which can be solved with standard numerical techniques. Xiu and Karniadakis (2002) and Xiu and Karniadakis (2003) have further used general polynomial chaos approach which employs a broader family of trial bases based on the orthogonal polynomials from the Askey scheme. The polynomial chaos approach combines the simplicity of Monte Carlo methods and stochastic Galerkin methods. However stochastic Galerkin procedure can be very challenging when the governing stochastic equations take complicated form as the derivation of the explicit equations for the polynomial chaos coefficients can be very difficult, if not impossible. Moreover, Galerkin projection formulation for the nonlinear froward simulator results in a system of coupled ordinary or partial differential equations. Hence, for complex physical system the numerical methods to solve the resulting deterministic equations can also be very computationally expensive. An alternative to stochastic Galerkin approach to the Bayesian inverse problem is the stochastic collocation approach (see Xiu and Hesthaven (2005), Xiu (2007), Marzouk and Xiu (2009) and Li and Zhang (2007)), which requires a finite number of uncoupled deterministic simulations of the forward model. The stochastic collocation method treat the forward

simulator as 'black box', so no reformulation is required for the governing equations of the forward model and the method can deal with highly non-linear problems that are challenging , if not impossible, to handle with stochastic Galerkin methods. However, the Stochastic collocation method imposes the requirement that the estimates of the model output are exact at a set of selected collocation points, thus making the residuals at those points exactly equal to zero. The unknown coefficients of the polynomial chaos is estimated by equating model outputs and the corresponding Hermite polynomial expansion at a set of collocation points or 'sampling points' in the parameter space called the nodes. The key issue in stochastic collocation method is the choice of sampling points. If the random input parameter is of one dimension then the stochastic collocation approach approximate the forward simulator using Lagrange interpolation of the sampling points and Gaussian quadrature formula is used to find the approximate solution of the polynomial chaos coefficients. Although quadrature is well developed for univariate case, less is known about the multivariate case. In our problem we deal with high dimension spatial input field which can be transformed through K-L expansion, Fourier series expansion or Discrete Cosine transform to multivariate random input parameters. The univariate quadrature scheme can be easily extended for multi dimensions by constructing a simple tensor product space. However, this is subject to the curse of dimensionality since the number of collocation points in a tensor grid grows exponentially fast in the number of input random variables and hence we have the same problem of running the simulator thousands of times to obtain the nodes of the collocation method.

Another very popular approach based on emulation of the simulators output has been developed that offers substantial efficiency gains over standard methods; for example see Sacks *et al.* (1989), Kennedy and O'Hagan (2001), O'Hagan (2006) and Conti *et al.* (2009). An emulator is a statistical representation of the simulator

that is constructed using a few training sample of simulator runs. Uncertainty and sensitivity analyses can be tackled using the emulator as the emulator runs essentially instantaneously as shown in Oakley and O'Hagan (2004). An emulator not only estimates the simulator output but also provides an entire probability distribution of the estimator. In practice emulator is an statistical approximation to the simulator which is built using statistical methods. Given a set of training runs of the forward model for some suitably chosen inputs, we treat both the inputs and the outputs as data with which we are to estimate the unknown function that maps inputs to the outputs. The emulator approach differs from the stochastic collocation approach, which also uses training samples using computer simulation, because the emulator being a statistical approximation not only approximate the forward simulator but it also quantifies the code uncertainty of the simulator. The efficiency gains arise because it is usually possible to emulate the simulator output to a high degree of precision using only a few hundred runs of the simulator. The key point in all the above mentioned articles on emulators is that the forward simulator is considered as an unknown function which is modeled by Gaussian processes regression or by other form of regression functions. The input model parameters of the simulator is either considered to be known or sometimes considered to be an unknown parameter which is calibrated using data from computer simulations. In this chapter we consider the spatial inverse problem where the input of the physical process contains a high dimension unknown spatial field. The available observations in the spatial process may be irregularly spaced and very sparse. We propose to use an emulator based on a Bayesian approach to multivariate adaptive regression spline (BMARS), as introduced in Denison *et al.* (1998), where the regressors are transformed parameters of the spatial field. In many physical process the input model parameters are of two kinds, one set of parameters are always completely known and the other set of parameters

are unknown which we want to calibrate. In this chapter we consider the case where the unknown input model parameter is a spatial process. Here we want to infer about the uncertainties of the input spatial process $Y(s,\omega), (\omega \in \Omega, s \in D \subset \mathcal{R}^2)$ given a limited number of observations on the spatial fields, a coarse scale representation of the spatial field and the observed output data. Suppose in addition, we have a number of independent simulator runs of the forward model, where the specified model inputs parameters, the spatial field and the corresponding outputs are completely known. We want to approximate the forward simulator by using BMARS emulators based on these simulated and real data. Difficulty arises while implementing BMARS because here the regressors consists of a high dimension spatial field. So at first we propose to use discrete cosine transformation on the spatial field where the original process $Y(s,\omega)$ is represented by a low dimensional parameterization. The finite dimensional transformed DCT coefficients and the other known input parameters are then used as the regressors while fitting the BMARS for the unknown forward simulator. To build the BMARS emulator, the input for simulation data were generated using the conditional Gaussian distribution and Latin hypercube design. Further the validation of the computer code is done by fitting the BMARS model on a training set data and applying the fitted model on the test data. Bayesian hierarchical method is used to model the posterior distribution of the unknown model parameters given the data. The hybrid sampling method, which is combination of reversible jump MCMC method, Metropolis Hastings method and Gibbs sampling method, is used to sample from the posterior. The BMARS based emulator method is very flexible, where the basis terms of the regression function are not fixed. The basis terms in the model are adaptively chosen by the data itself making the model more flexible than the collocation method and polynomial chaos method.

There are various examples of spatial inverse problems such as in the areas of

ground water flow where the unknown input parameters is the hydrolic conductivity, weather forecasting where the unknown input model parameter are initial velocity and height field, chemical kinetics and reservoir characterization. For the definiteness of our problem, in this chapter we will consider examples in the area of reservoir characterization where the the single most influential input is the unknown permeability spatial field. The other model input parameters, such as porosity, the pore volume injected at the injectors are considered to be known. The output observations are fractional flow or water-cut data which is the fraction of water produced in relation to the total production rate in a two phase oil-water flow reservoir. We assume that the values of the fine-scale permeabilities are known at the wells and the permeability data in a coarse-scale are available. We also have data from some independent simulation run of the forward model. We use the simulated and the available real data to infer about the unknown permeability spatial field by sampling from the posterior distribution of the model parameters given the data.

Our numerical results illustrate that the proposed BMARS model based on the transformed DCT coefficient regressors can predict the output from the simulated test data adequately. The Bayesian model can also predict the unknown permeability field and its uncertainty very well. The computation efficiency in terms of CPU time for the emulator based method are also shown to be very high when compared to the simulation based methods.

The chapter is organized as follows. In the Subsection IV.1 we discuss the DCT parametrization of the spatial field. In Subsection IV.2, we formulate the Bayesian Hierarchical model where we discuss how we can use the BMARS model to build the emulator using simulation runs. In Subsection IV.3, we discuss the sampling procedure used to sample from the posterior. In Subsection IV.4, numerical results are presented for simulated as well as real data set.

IV.1. Parametrization of the Spatial Field Using Discrete Cosine Transform (DCT)

The Karhunen-Loève (K-L) expansion discussed in the last chapter is a classical way for deriving low-dimensional parameterization of a high dimension spatial field. It is shown that for a known covariance structure of a spatial field the K-L expansion provides optimal compression in terms of the mean square(MSE)among all linear transformation of the spatial field. However, one of the major disadvantage of the K-L expansion is the requirement of decomposition of a high dimension covariance matrix which makes the problem computationally very expensive. Additionally building an emulator based on the K-L transform of the spatial field makes the problem more complicated as the basis functions of the K-L expansion depends on the type of the correlation function and the parameters of the correction function of the spatial field. Here we consider a much more inexpensive form of parameterization of a spatial field based on discrete cosine transform (DCT) as described in Jafarpour and McLaughlin (2007). The DCT approach uses a set basis vectors that are predefined and do not depend on the covariance structure of the spatial field. More precisely, the transformation kernels used in the DCT are real cosine functions. Hence we can use the transformed DCT co-efficients as the regressors of the BMARS emulator. The DCT approach also provides an faster alternative to the KLT approach and also requires fewer assumptions Ahmed *et al.* (1974). Moreover, for first order stationary Markov processes, DCT bases asymptotically converge to KLT bases. The computational order of DCT is $O(Nlog_2N)$ (see Brigham (1988), Jain (1989), Feig and Winpgrad (1992) and Narasima and Peterson (1978)) where as the computational order of KLT is $O(N^3)$ using singular value decomposition of the covariance matrix (see Jain (1989)). More details of the formulation, derivation, and properties of DCT and its comparison to KLT and other Fourier-type transformations can be found in

Rao and Yip (1990) and Ersoy (1994). In our simulation model the DCT basis vectors for a spatial field are computed only once and stored for use as they are pre-specified and do not depend on the data. The DCT basis vectors are also orthogonal which facilitates the computation of the inverse transform very easily. The transformation is also separable so that it can be applied to multi-dimensional spatial field one dimension at a time (see Gonzalez (2002) and Rao and Yip (1990)) resulting in substantial reduction in computation time.

Using DCT transformation an one dimensional field $Y(s)$ of length N can be written as

$$Y(s) = \sum_{k=0}^{N-1} \alpha_k \theta_k cos\left[\frac{\pi(2s+1)k}{2N}\right], \tag{4.1}$$

where,

$$\theta_k = \alpha_k \sum_s Y(s).cos\left[\frac{\pi(2s+1)k}{2N}\right], 0 \leq k \leq N-1, \tag{4.2}$$

$$\alpha_k = \sqrt{2},\ k = 1 : N-1\ \alpha_0 = \sqrt{\frac{2}{N}}. \tag{4.3}$$

Extensions of 4.1 to higher dimensions can be done by using the separability property of DCT, i.e. by applying the 1D transform in each direction. Using DCT transformation a two dimensional spatial field $Y(s)$ ($s = (s_x, s_y)$) of dimension $N \times N$, can be written as

$$Y(s_x, s_y) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \alpha_i \alpha_j \theta_{ij} cos\left[\frac{\pi(2s_x+1)i}{2N}\right] cos\left[\frac{\pi(2s_y+1)j}{2N}\right], \tag{4.4}$$

where,

$$\theta_{ij} = \alpha_i \alpha_j \sum_{s_x} \sum_{s_y} Y(s_x, s_y).cos\left[\frac{\pi(2s_x+1)i}{2N}\right] cos\left[\frac{\pi(2s_y+1)j}{2N}\right], 0 \leq i, j \leq N-1, \tag{4.5}$$

It is clear from the above equation that it is derived by multiplying the horizontal one dimensional basis function withe the vertical one dimensional basis function.

By using DCT the spatial image could be transformed and truncated to few DCT coefficients while preserving the large-scale continuity of the spatial property (see Figure 29). As described in Jafarpour and McLaughlin (2007), generally the large DCT coefficients are concentrated on the top left corner of the transformed space. This clustering of coefficients generally corresponds to the modes with large scale variations in the horizontal, vertical, and diagonal directions. If the dominant features of the spatial field are expected to be vertically oriented more coefficients should be retained from the left side of the transformed DCT coefficients. If the dominant features are horizontally oriented more coefficients should be retained from the top of the DCT coefficient array. When suffiecient prior information is available, proper coefficients can be selected more systematically. In our spatial inverse problem the spatial field and its DCT coefficients are completely unknown, so it is almost impossible to identify the retained modes by ordering the coefficients. In the absence of any prior information, a reasonable orientation insensitive alternative is to retain modes associated with coefficients inside a diagonally symmetric triangle or a square in the top left corner of the transformed space (see Jafarpour and McLaughlin (2007)). So we can write the spatial process $Y(s, \omega), (\omega \in \Omega, s = (s_x, s_y) \in D \subset R^2)$ as a linear combination of the truncated transformed DCT coefficients as

$$Y(s, w) = B_c(s)\theta(w), \tag{4.6}$$

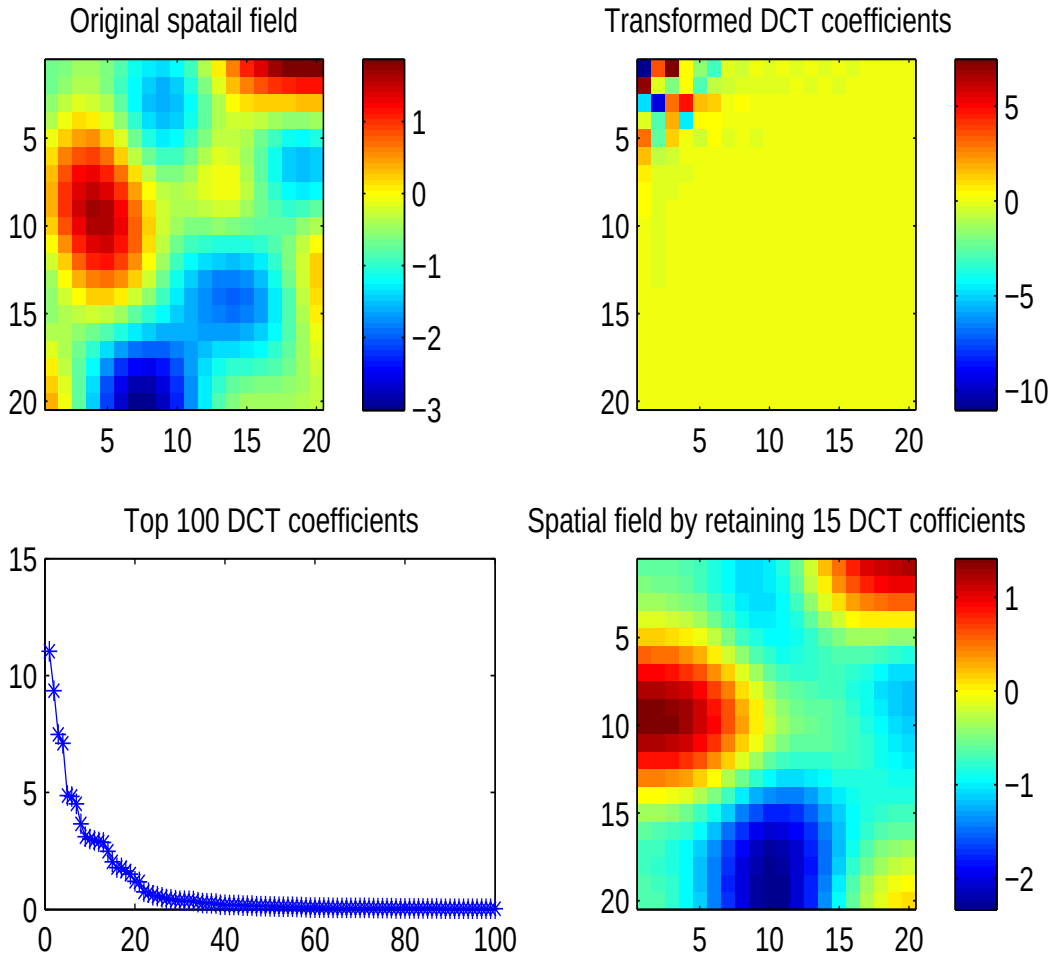

Fig. 29. A spatial field, the corresponding DCT coefficients and the spatial field obtained by the inverse DCT transform.

where, θ is the vector of the truncated DCT coefficients and B_c is the matrix of the DCT basis functions. Note that here θ may be of dimension $k << N^2$, the dimension of the original spatial field. So given the truncated DCT coefficients θ we can completely specify the spatial field by the corresponding inverse DCT transformation 4.6 as the DCT basis functions are predetermined and fixed.

IV.2. The Bayesian Hierarchical Model

For the spatial inverse problem various mathematical models are used to describe some physical process over a spatial field. Often the mathematical models are highly complex so that the computer code is very expensive in terms of CPU time required for a single run. Moreover, the computer model often have a number of context specific inputs that define a particular situation in which the model to be used. In this chapter we consider the case where the output depends on two groups of inputs to the the computer model. One of the input is a unknown spatial field which we want to learn about using the output observations. The other group comprises of all other model parameters which are completely known for each of the output observations. We want to learn about the unknown spatial field and its uncertainties. The method is described for a unknown input parameter vector in Kennedy and O'Hagan (2001) and is called Bayesian calibration. In this chapter we want to extend the idea of Bayesian calibration where the unknown input is a high dimension spatial field. Here we consider both the uncertainty accounted for the input unknown spatial field together with the uncertainties of the computer codes. We first build an emulator for the unknown function that maps the output to the inputs, using data from some simulation runs and the observed output data. The output of the computer model or the simulator is denoted as $z_i^s, i = 1, 2, \ldots n_s$. The corresponding known input model parameters are denoted as $x_i^s, i = 1, 2, \ldots n_s$. The calibration inputs for the

spatial field are obtained by a Latin hypercube sampling of the DCT coefficients (as described in details in the next subsection), say $x_i^t, i = 1, 2, \ldots n_s$ and then using the corresponding inverse DCT transformation. The model for the simulated data is given by

$$z_i^s = \eta(x_i^s, x_i^t) + \epsilon_i^s, i = 1, 2, \ldots n_s. \tag{4.7}$$

Similarly, we also have additional n_r , ($n_r << n_s$) real data from the output $z_i^r, i = 1, 2 \ldots n_r$. The corresponding known input variables are denoted as $x_i^r, i = 1, 2, \ldots n_r$ and the corresponding unknown spatial field is parameterized by the unknown DCT coefficients θ. So the model for the observed data is given by

$$z_i^r = \eta(x_i^r, \theta) + \epsilon_i^r, i = 1, 2, \ldots n_r. \tag{4.8}$$

We assume the model error for the simulator or computer code follows $N(0, \sigma_z^2)$ and the observational error for the output observations as $N(0, \tau_z \sigma_z^2)$.

In additional to the output observations we assume that a coarse scale data is available for the unknown spatial field denoted by y_c and a few sparse fine scale observations are also available for the spatial field denoted by y_o

The hierarchical model is given by

$$\begin{aligned} P(\theta, \sigma_z^2, \tau_z | x_{1:n_s}^t, \quad x_{1:n_r}^r \quad & , x_{1:n_s}^s, z_{1:n_r}^r, z_{1:n_s}^s, y_c, y_o) \\ \propto \quad & P(\theta, \sigma_z^2, \tau_z, x_{1:n_r}^r, x_{1:n_s}^s, z_{1:n_r}^r, z_{1:n_s}^s, y_c, y_o | x_{1:n_s}^t) \\ = \quad & P(z_{1:n_r}^r, z_{1:n_s}^s | \theta, \sigma_z^2, \tau_z, x_{1:n_r}^r, x_{1:n_s}^s, x_{1:n_s}^t) \\ & \times P(y_c|\theta) P(y_o|\theta) P(\theta) P(\tau_z) P(\sigma_z^2). \end{aligned} \tag{4.9}$$

The first term in the hierarchical model 4.9 contains the unknown function η which maps the outputs to the inputs and signifies the probability distribution of the output given the inputs. The second term contains an upscaling operator as

described before and denotes the probability distribution of the coarse scale data given the spatial field. The third term denotes the probability distribution of the observed fine scale spatial data given the DCT coefficients. In the next subsections first we will describe about the designs used for sampling the unknown input spatial field for the n_s simulation run, then we will elaborate on how we model each of the terms in the Bayesian hierarchical model.

IV.2.1. Design of the Simulation Experiments

One of the very important issue in building emulator is the method of choosing the input configurations at which we are to run the simulator model to get the training data. The main objective of choosing the input design configuration is to learn about the unknown function that maps the input to the output over well spaced input points that cover the wide region of interest as close as possible. In our problem the unknown inputs of the forward model is a random spatial field. The usual way to create multiple realizations of spatial random field is to draw a simple random sample from a Gaussian process with a proper correlation structure. Since the computer simulator is very expensive for a complicated physical process, the number of samples of the spatial field has to be kept small for practical reason. In that case more accurate assessment of the emulator can be obtained when more efficient sampling method such as Latin hypercube sampling (see McKay *et al.* (1979) and Ross (1990)) are used. In McKay *et al.* (1979) it is shown that the estimate of the unknown output using Latin hypercube samples have lesser variance than that of using simple random samples. In our problem we have applied two different approaches of sampling the input random spatial field. In the first approach we consider the case where enough data on the fine scale spatial field, y_o, are available to fit a semi-variogram model using weighted least squares (see Cressi (1993)). In this case we use Latin hypercube

sampling technique to sample from the conditional Gaussian distribution of the spatial process in the unknown spatial locations given the spatial field at the known locations as described in Pebesma and Heuvelink (1999). In the second approach we consider the case where very few fine scale realization of the fine-scale spatial field is available, but we have data from a relatively coarser scale, y_c. In this case we first transform the available coarse-scale data to a fine-scale data by replacing every element in the fine-scale of a coarse-block with the corresponding coarse scale value. Then we apply DCT transformation of the obtained fine scale data and truncate the DCT terms up to a desired degree of accuracy. Suppose the transformed DCT coefficients are θ_{obs}. First we obtain a large number samples of the DCT coefficients, say N_s samples, using Latin hypercube sampling from a multivariate normal distribution with mean θ_{obs} and variance $\gamma\mathbf{I}$, where γ is large. Latin Hypercube sampling is a stratified random sampling technique in which a sample of a fixed size from multiple variables is drawn such that for each individual variable the sample is (marginally) maximally stratified. Given a simple random sample from the target multivariate distribution, a Latin hypercube sample from the same distribution is obtained by slightly shifting the sample elements (see Stein (1987)). After N_s samples are obtained we choose n_s samples from them having the largest minimum distance between the set of points. The corresponding n_s input spatial fields are obtained from the inverse transformation of the sampled DCT coefficient.

IV.2.2. Modeling the Likelihood Using Bayesian MARS Emulators

In this subsection we describe how we can use Bayesian approach to multivariate adaptive regression spline (BMARS) to approximate the forward simulator F as describe by the models 4.7 and 4.8. We assume that there are k_1 input variables that are completely known for each out put, so that $x_i^s = (x_{i1}^s, x_{i1}^s \dots x_{ik_1}^s)', i = 1, 2, \dots n_s$

and $x_i^r = (x_{i1}^r, x_{i1}^r \dots x_{ik_1}^r)', i = 1, 2, \dots n_r$. Suppose we retain k_2 coefficients in the DCT transformation, i.e $x_i^t = (x_{i1}^t, x_{i2}^t \dots x_{ik_2}^t)', i = 1, 2, \dots n_s$.

Combining the simulated data and the observed data we can write the model 4.7and 4.8 as

$$Z = \eta(X) + \varepsilon, \tag{4.10}$$

where, $X = \begin{pmatrix} x_{11}^s & x_{12}^s & \dots & x_{1k_1}^s & x_{11}^t & x_{12}^t & \dots & x_{1k_2}^t \\ x_{21}^s & x_{12}^s & \dots & x_{2k_1}^s & x_{21}^t & x_{22}^t & \dots & x_{2k_2}^t \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n_s1}^s & x_{n_s2}^s & \dots & x_{n_sk_1}^s & x_{n_s1}^t & x_{n_s2}^t & \dots & x_{n_sk_2}^t \\ x_{11}^r & x_{12}^r & \dots & x_{1k_1}^r & \theta_1 & \theta_2 & \dots & \theta_{k_2} \\ x_{21}^r & x_{22}^r & \dots & x_{2k_1}^r & \theta_1 & \theta_2 & \dots & \theta_{k_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n_r1}^r & x_{n_r2}^r & \dots & x_{n_rk_1}^r & \theta_1 & \theta_2 & \dots & \theta_{k_2} \end{pmatrix} = \begin{pmatrix} X_s & X_t \\ X_r & X_\theta \end{pmatrix}$,(say)

$Z = \left(z_1^s, z_2^s \dots z_{n_s}^s, z_1^r, z_2^r, \dots z_{n_r}^r\right)' = (Z_s', Z_r')'$, (say)

$\varepsilon = \left(\epsilon_1^s, \epsilon_2^s \dots \epsilon_{n_s}^s, \epsilon_1^r \dots \epsilon_{n_r}^r\right)' = (\varepsilon_s', \varepsilon_r')'$, (say).

The multivariate adaptive regression splines (MARS)Friedman (1991) is a regression method where the unknown simulator or black box is approximated by a regression function given by

$$\eta(x) = \sum_{i=1}^{m} \beta_i B_i(x), \tag{4.11}$$

where, the basis function $B_i(x)$ is given by

$$B_i(x) = \begin{cases} 1 & i = 1 \\ \prod_{j=1}^{J_i} \left[s_{ji}.(x_{\nu(ji)} - t_{ji})\right]_+, & i = 2, 3, .. \end{cases} \tag{4.12}$$

where $(.)_+ = max(0, .)$, J_i is the degree of the interaction of basis B_i, the s_{ji}, which we shall call the sign indicators, equal ± 1, the $\nu(j, i)$ give the index of the predictor variable which is being split on the t_{ji} (known as knot points) give the position of

the splits. The $\nu(j,.)$ $(j = 1 \dots J)$ are constrained to be distinct so each predictor only appears once in each interaction term. Frequently some maximum order of interaction I is assigned to the model such that $J_i \leq I(i = 1 \dots m)$. We let $T_i \in \{1, 2, \dots N_I\}(N_I = \sum_{i=1}^{I} \binom{k1+k2}{i})$ denote the type of basis function B_i thus T_i, in effect, just tells us which predictor variables we are splitting on, i.e. what the values of $\nu(1,i), \dots, \nu(J_i,i)$ are. We denote all the parameters in the model as $(m, c^m, \beta^{\mathbf{m}})$. Here $c^m = (\mathcal{V}_1, \dots, \mathcal{V}_m)$ where each $\mathcal{V}_i$ is the $(1+2J_i)$dimensional vector $(T_i, t_{1,i}, s_{1,i}, \dots, t_{J,i}, s_{J,i})$ which corresponds to basis function B_i. We assume

$$\epsilon_i^s \overset{\text{iid}}{\sim} N(0, \sigma_z^2), i = 1 : n_s, \tag{4.13}$$

and

$$\epsilon_i^r \overset{\text{iid}}{\sim} N(0, \tau_z \sigma_z^2), 1 = 1 : n_r. \tag{4.14}$$

So, $P(Z|X, m, c^m, \beta^{\mathbf{m}}, \sigma_z^2,)$ is the pdf of the Multivariate Normal Distribution with mean $\sum_{i=1}^{m} \beta_i B_i(X)$ and variance $\Sigma = \sigma_z^2 \begin{pmatrix} I_{n_s} & \mathbf{0} \\ \mathbf{0} & \tau_z I_{n_r} \end{pmatrix}$.

Bayesian multivariate adaptive regression spline assigns a prior distribution to every unknown parameters in the model. We assume vague, but proper, prior for σ_z^2 , i.e $\sigma_z^2 \sim Inverse\ Gamma(a_z, b_z)$. Similarly we assume $\tau_z \sim Inverse\ Gamma(a_\tau, b_\tau)$ T_i are assumed to be uniformly distributed on $\{1, 2, \dots N\}$. The sign indicators s_{ji} and knot points t_{ji} are also assumed to be uniformly distributed on the sets $\{1, -1\}$ and $\{1, 2, \dots n\}$ respectively. We use another vague, but proper prior, for the coefficients of the basis functions, i.e we assume the $\beta_i \sim N(0, \alpha\sigma_z^2)$, where α is very large. The prior distribution of m is taken to be truncated Poisson distribution with parameter λ, truncated at m_{max}. The prior distribution for θ is given by:
$\theta|\sigma_\theta^2 \sim MVN(0, \sigma_\theta^2 I)$ and $\sigma_\theta^2 \sim InverseGamma(a_o, b_o)$. After integrating out σ_θ^2 we

obtain the marginal prior distribution as

$$P(\theta) \propto \frac{\Gamma(a_o + k_2/2)}{[b_o + \frac{1}{2}\theta'\theta]^{(a_o+k_2/2)}}. \tag{4.15}$$

Applying the BMARS model the posterior distribution of the parameters is given by:

$$\begin{aligned}
\pi(\theta, \sigma_z^2, \tau_z, m, \beta^{\mathbf{m}}, c^m) &= P(\theta, \sigma_z^2, \tau_z, m, \beta^{\mathbf{m}}, c^m | X_t, X_r, X_s, Z, y_c, y_o) \\
&\propto P(\theta, \sigma_z^2, \tau_z, m, \beta^{\mathbf{m}}, c^m, , Z, y_c, y_o | X_t, X_r, X_s) \\
&= P(Z|X, m, \beta^{\mathbf{m}}, c^m, \sigma_z^2, \tau_z) P(y_c|\theta) P(y_o|\theta) \\
&\times P(\theta) P(\tau_z) P(\sigma_z^2) P(\beta^{\mathbf{m}}|m) P(c^m|m) P(m). \qquad (4.16)
\end{aligned}$$

IV.2.3. Modeling the Coarse Scale Data

In many cases the coarse scale data are readily available which contain important information to reduce the uncertainty in estimation of the fine-scale spatial field. Upscaling procedure is a way to link the coarse and the fine scale data. The simplest way to think about the upscaling procedure in the spatial domain is the use of spatial block averages of the fine-scale spatial data to obtain the coarse scale data. We need to modify this averaging idea in a way so that the forward equations (and the corresponding boundary conditions) remained valid in this upscaling scheme. The main idea of our approach is to upscale the spatial field Y on the coarse-grid, then solve the original system on the coarse-grid with upscaled spatial field. The main theme of the procedure is that given a fine scale spatial field Y, we can use a operator L (it can be averaging or more complicated integrations with boundary conditions) so that the coarse data y_c can be expressed as $y_c = L(Y) + \epsilon_c$, where ϵ_c is a random error term which explains the variations from deterministic upscaling procedures. As we have parameterized the spatial field Y using the DCT, the final equation is given

as

$$y_c = L_c(\theta) + \epsilon_c, \tag{4.17}$$

where L_c can be looked upon as an operator whose input is the fine-scale spatial field or the transformed parameters of the model θ and output is the coarse-scale spatial field. We assume that the error ϵ_c follows a multivariate normal distribution with mean 0 and covariance $\sigma_c^2 I$. i.e $y_c|\theta, \sigma_c^2 \sim MVN(L_c(\theta), \sigma_c^2 I)$. We assume the prior distribution of σ_c^2 as $\sigma_c^2 \sim InverseGamma(a_c, b_c)$. Furthermore, after integrating out σ_c^2 we obtain the marginal distribution as

$$P(y_c|\theta) \propto \frac{\Gamma(a_c + N^*/2)}{[b_c + \frac{1}{2}||(y_c - L_c(\theta)||^2]^{(a_c+N^*/2)}}, \tag{4.18}$$

where N^* is he number of observations of the coarse-scale permeability field.

IV.2.4. Modeling the Observed Fine Scale Data

The fine scale observations are obtained at some locations of the spatial field and we specify a model for $P(y_o|\theta)$ as

$$y_o = y_p + \epsilon_k, \tag{4.19}$$

where y_p is the the fine scale spatial field at the given locations of y_o obtained from the inverse DCT transformation of θ as described. ϵ_k is the model error for the DCT approximation. We assume ϵ_k follows a multivariate normal distribution with mean 0 and covariance $\sigma_k^2 I$. i.e, $y_o|\theta, \sigma_k^2 \sim MVN(y_p, \sigma_k^2)$. The prior for σ_k^2 is assumed to be $\sigma_k^2 \sim InverseGamma(a_k, b_k)$. After integrating out σ_k^2 we obtain

$$P(y_o|\theta) \propto \frac{\Gamma(a_k + N_{obs}/2)}{[b_k + \frac{1}{2}(y_o - y_p)'(y_o - y_p)]^{(a_k+N_{obs}/2)}}, \tag{4.20}$$

where N_{obs} is he number of observations of the fine-scale spatial field.

IV.3. Sampling from the Posterior

Our aim is to simulate from the joint posterior distribution $\pi(\theta, \sigma_z^2, \tau_z, m, \beta^{\mathbf{m}}, c^m)$ We use a hybrid sampling method which is a combination of Gibbs sampling and metropolis hastings algorithm to sample from the posterior. Note that our method deals with multiple parameter subspaces of different dimension so we use reversible jump MCMC algorithm as discussed in Green (1995). An algorithm of the sampling procedure from the posterior distribution is given below.

IV.3.1. Hybrid Sampling Algorithm

- Step 1. First we want to sample from the joint conditional distribution of $P(m, c^m, \beta^{\mathbf{m}}, \sigma_z^2 | \theta, \tau_z, X_t, X_r, X_s, Z, y_c, y_o)$. In order to sample from this joint distribution we do the following three sub-steps.

    1. The marginal likelihood of $P(m, c^m | \theta, \tau_z, X_t, X_r, X_s, Z, y_c, y_o)$ can be computed up to a constant of proportionality by marginalizing over σ_z^2 and β. We sample from this marginal distribution using reversible jump MCMC technique where at each step we use one of the following types of moves: (a)the addition of a basis function (BIRTH); (b) deletion of a basis function (DEATH); (c) a change in a knot location (CHANGE). The addition of basis function, move type (a) is carried out by choosing uniformly a type of basis function, say T_i to add to the model. Then a knot location and sign indicator for each of the J_i factors in the new basis is chosen uniformly. In move type (b) the deletion of a basis function is done in such a way as to make the jump step reversible. This is done by removing an uniformly chosen basis function from the present(except the constant basis function).

2. $P(\beta|m, c^m, \tau_z, \theta, X_t, X_r, X_s, Z, y_c, y_o)$ can be calculated completely after integrating over σ_z^2 , so we use a Gibbs sampling step to sample from this marginal distribution.

3. $P(\sigma_z^2|m, c^m, \beta^{\mathbf{m}}, \tau_z, \theta, X_t, X_r, X_s, Z, y_c, y_o)$ can be calculated completely, so we use a Gibbs sampling step to sample from this marginal distribution.

- Step 2. $P(\tau_z|m, c^m, \beta^{\mathbf{m}}, \sigma_z^2, \theta, X_t, X_r, X_s, Z, y_c, y_o)$ can be calculated completely, so we use a Gibbs sampling step to sample from this marginal distribution.

- Step 3. $P(\theta|m, \beta^{\mathbf{m}}, c^m, \sigma_z^2, \tau_z, X_t, X_r, X_s, Z, , y_c, y_o)$ can be calculated up to a constant of proportionality , so we use Metropolis Hastings step to sample from this conditional distribution.

Now we will describe the details about he conditional and marginal distributions used in the algorithm IV.3.1.

Step 1.

$$\begin{aligned} P(m, c^m|\theta, \tau_z, X_t, X_r, X_s, Z, y_c, y_o) &\propto \int_{\beta^m}\int_{\sigma_z^2} P(Z|X, \sigma_z^2, \tau_z, m, c^m, \beta^{\mathbf{m}}) \\ &\times P(c^m|m)P(\beta^m|m)P(m)P(\sigma_z^2)d\sigma_z^2 d\beta^m \\ &= \pi_1(m, c^m|\theta, \tau_z), (\text{say}). \end{aligned} \tag{4.21}$$

It can be shown that

$$\begin{aligned} log(\pi_1) &= C_1 - (\frac{n}{2} + a_z)log(d) - \frac{m}{2}\tau_z + (m-1)log(\lambda) - log(p!) - log(\frac{p(p+3)}{2}) \\ &\quad -\frac{m}{2}log(\alpha) - log\left(\Gamma\left(\frac{n}{2} + a_z + \frac{m}{2}\right)\right)\sum_{j=1}^{m} J_j log(2n), \end{aligned} \tag{4.22}$$

where,

$$d = 2b_z + Z'\Sigma^{-1}Z - (\mathbf{B}'\Sigma^{-1}Z)\Sigma_k^{-1}(\mathbf{B}'\Sigma^{-1}Z),\ \Sigma_k = \left[\mathbf{B}'\Sigma^{-1}\mathbf{B} + I/\alpha\right]^{-1},$$

$$p = k_1 + k_2,\ C_1 = \frac{1}{2} log|\Sigma_k| \text{ and } \mathbf{B} = (B_1, B_2, \ldots, B_m). \quad (4.23)$$

So, the acceptance probability of reversible jump MCMC step as described in first part of step one in the algorithm IV.3.1 can be written as

$$\alpha = min\left(1, \frac{\pi_1(m', c'^{m'}|\tau_z, \theta) S((m', c'^{m'}) \to (m, c^m))}{\pi_1(m, c^m|\tau_z, \theta) S((m, c^m) \to (m', c'^{m'}))}\right), \quad (4.24)$$

where (m, c^m) denotes the current model parameters and $(m', c'^{m'})$ denotes the proposed model parameters. At each iteration we either have a BIRTH move or DEATH move or CHANGE move step with probability b_m, d_m and ρ_m respectively. The proposal ratio for a BIRTH move is given by

$$\begin{aligned} \frac{S((m', c'^{m'}) \to (m, c^m))}{S((m, c^m) \to (m, c'^{m'}))} &= \frac{P\left(\text{propose death}(m+1, c'^{m+1}) \to (m, c^m)\right)}{P\left(\text{propose birth}(m, c^m) \to (m+1, c'^{m+1})\right)} \\ &= \frac{d_{m+1}/m}{2b_m/[p(p+3)(2n)^{J_{m+1}}]}. \end{aligned} \quad (4.25)$$

The proposal ratio for a DEATH move is given by

$$\begin{aligned} \frac{S((m', c'^{m'}) \to (m, c^m))}{S((m, c^m) \to (m, c'^{m'}))} &= \frac{P\left(\text{propose birth}(m-1, c'^{m-1}) \to (m, c^m)\right)}{P\left(\text{propose death}(m, c^m) \to (m-1, c'^{m-1})\right)} \\ &= \frac{2b_{m-1}/[p(p+3)(2n)^{J_{m-1}}]}{d_m/(m-1)}. \end{aligned} \quad (4.26)$$

The proposal ratio for a CHANGE move is always 1. Note that here $b_m + d_m + \rho_m = 1,\ \forall m$. In particular we take $b_m = d_m = \rho_m = \frac{1}{3}\ \forall m = 2, 3, \ldots m_{max}$, $b_1 = 1, d_1 = \rho_1 = 0$, $b_{m_{max}} = 0, d_{m_{max}} = 1$ and $\rho_{m_{max}} = 0$.

Similarly the Gibbs step in the second part of step one in the algorithm IV.3.1 is carried out by generating samples from the marginal conditional distribution of β^m given $m, c^m, \tau_z, \theta, x^t_{1:n_s}, x_{1:n}, Z, , y_c, y_o$ given by

$$\beta^m|m, c^m, \tau_z, \theta, X_t, X_r, X_s, Z, y_c, y_o \sim t_{n+2a_z}(\Sigma_k[\mathbf{B}'\Sigma^{-1}Z], \frac{d}{n+2a_z}\Sigma_k). \quad (4.27)$$

In the third part of step one in the algorithm IV.3.1 the Gibbs step is done by simulating from the conditional distribution of σ_z^2 given $m, c^m, \beta^{\mathbf{m}}, \tau_z, \theta, X_t, X_r, X_s, Z, y_c, y_o$, which is given by

$$\sigma_z^2 | m, c^m, \beta^{\mathbf{m}}, \tau_z, \theta, X_t, X_r, X_s, Z, y_c, y_o \sim IG(\delta_{z1}, \delta_{z2}), \tag{4.28}$$

where

$$\delta_{z1} = a_z + \frac{m+n}{2}, \delta_{z2} = b_z + \frac{(Z - \mathbf{B}\beta)'\Sigma^{-1}(Z - \mathbf{B}\beta)}{2} + \frac{\beta^2}{2\alpha^2}. \tag{4.29}$$

Step 2. For a given m, c^m we write the matrix of the basis functions defined on BMARS as:

$$\mathbf{B} = (B_1, B_2, \ldots, B_m) = \begin{pmatrix} B_{11} & B_{12} & \ldots & B_{1m} \\ \vdots & \vdots & \vdots & \vdots \\ B_{n_s 1} & B_{n_s 2} & \ldots & B_{n_s m} \\ B_{n_s+11} & B_{n_s+12} & \ldots & B_{n_s+1 m} \\ \vdots & \vdots & \vdots & \vdots \\ B_{n1} & B_{n2} & \ldots & B_{n_s m} \end{pmatrix} = \begin{pmatrix} \mathbf{B_s} \\ \mathbf{B_r} \end{pmatrix}, \text{(say)}.$$

The Gibbs sampling in step two of the algorithm IV.3.1 is carried out by sampling from the conditional distribution of τ_z given $m, c^m, \beta^{\mathbf{m}}, \sigma_z^2, \theta, X_t, X_r, X_s, Z, y_c, y_o$, which is given by

$$\tau_z | m, c^m, \beta^{\mathbf{m}}, \sigma_z^2, \theta, X_t, X_r, X_s, Z, y_c, y_o \sim IG(\delta_{\tau 1}, \delta_{\tau 2}), \tag{4.30}$$

where

$$\delta_{\tau 1} = \frac{n_r}{2} + a_\tau, \delta_{\tau 2} = \frac{(Z_r - \mathbf{B_r}\beta^m)'(Z_r - \mathbf{B_r}\beta^m)}{2\sigma_z^2} + b_\tau. \tag{4.31}$$

Step 3.

The conditional distribution of θ given $m, c^m, \beta^{\mathbf{m}}, \sigma_z^2, \tau_z, X_t, X_r, X_s, Z, y_c, y_o$ as used in step three of the algorithm is described below. Let us denote $P(\theta | m, c^m, \beta^{\mathbf{m}}, \sigma_z^2, \tau_z, X_t, X_r, X_s, Z, y_c, y_o)$ as $\pi_2(\theta | m, c^m, \beta^{\mathbf{m}}, \sigma_z^2, \tau_z)$, then

$$
\begin{aligned}
log(\pi_2) \quad &\propto \quad -(Z_r - \mathbf{B_r}\beta^m)'(Z_r - \mathbf{B_r}\beta^m)/(2\sigma_z^2\tau_z) \\
&- \quad (a_c + N^*/2)log\left(b_c + (y_c - L_c(\theta))'(y_c - L_c(\theta))/2\right) \\
&- \quad (a_k + N_{obs}/2)log\left(b_k + (y_o - y_p)'(y_o - y_p)/2\right) \\
&- \quad (a_o + k_2/2)log\left(b_o + \theta'\theta/2\right). \qquad (4.32)
\end{aligned}
$$

The metropolis hastings sampling in the third step of the algorithm IV.3.1 is carried out by first proposing a new parameter $\theta' = \theta + h\xi$, where ξ is a random variable and h is the jump size. We accept θ' with probability given by

$$\alpha = min\left(1, \frac{\pi_2(\theta')q(\theta'|\theta)}{\pi_2(\theta)q(\theta|\theta')}\right), \qquad (4.33)$$

where $q(\theta|\theta')$ is the proposal distribution of θ given θ'.

IV.4. Numerical Results

As described before the spatial inverse problems has application in many fields such as ground water flow, chemical kinetics, weather forecasting and reservoir characterization. For the definiteness of our problem we only consider examples from reservoir characterization. Subsurfaces, more specifically petroleum reservoirs are complex geological formations encompassing a wide range of physical and chemical heterogeneities. The goal of stochastic models is to characterize the different attributes of reservoirs such as permeability, porosity, fluid saturation, and etc. Here we concentrate on the single most influential attribute in reservoir characterization, the permeability spatial field denoted by k. As permeability takes positive values, hence we transform $Y = log(k)$ for our modeling convenience. The main available response is the fractional flow or the water-cut data which is the fraction of water produced in relation to

the total production rate in a two phase oil-water flow reservoir denoted by z. There are two kinds of inputs of the forward simulator: (i) a permeability spatial field Y which can be parametrized by DCT transformed coefficients θ and (ii) the pore volume injected. Given the inputs of the model the output or the water-cut observations can be obtained from the Darcys law which contains several partial differential equations. The goal of our inverse model is to learn about the known input spatial field given the output z, some coarse scale realization of the permeability field, y_c and some fine scale realization of the spatial field y_o. Here we focus on the problem where we can approximate the forward simulator by emulator based on a few simulation runs of the computer code for the forward simulator.

IV.4.1. Simulated Reservoir Example

In our first example we considered only simulated data from a reservoir characterization example. As discussed before the output is fraction flow or water-cut data and the two types of inputs are a 25×25 spatial field on a unit square and pore volume injected. Each grid on fine scale data is of 0.04×0.04 square unit. First we generate 100 samples of 15 DCT coefficients using Latin hypercube sampling with multivariate normal distribution. Each of these 15 DCT coefficients corresponds to a log permeability field obtained by the inverse DCT transformation of a matrix whose elements of the top left corner triangle are those 15 DCT coefficients, the remaining elements set to zero. Thus we have 100 simulated realizations of the input spatial field. The other input variable is pore volume injected at the injector wells. The simulated output is the fractional flow or water-cut data which is obtained by running the computer simulator of the forward model. For each simulated spatial field we have 50 outputs corresponding to 50 known input of second type (pore volume injected). An additional spatial log permeability field is obtained by simulating from a

Gaussian random field with known covariance structure which is treated as reference permeability field, we again use the same computer code to simulate the corresponding outputs. We treat this spatial field as unknown which we want to calibrate. Since no real data is used in the model so here $\tau_z = 1$. Before we do the calibration, we first build an emulator based on a portion of the simulated data, called the training data and see how our emulator performs on the test data. So we first we divide the 5000 simulated data set on inputs and outputs corresponding to 100 simulated spatial field into two parts. The first 4500 data set corresponding to 90 simulated spatial permeability field are called the training data set. The rest 500 data set corresponding to 10 permeability field are called test data set. First we built the BMARS emulators where the regressors are the DCT coefficients and pore volume injected and the response is the logit transformation of the watercut. Then we use the fitted model to predict the output for the test data. This process is called computer model validation. In Figure 30 we can see from the scatter plot of the mean of fitted output vs the simulated output that all the observations lies almost on the straight line through the origin. Also from the box plot of the predicted errors we can see the median of the errors is close to zero which explains that the BMARS emulators can predict the output very well. The simulated output, the corresponding fitted mean and the 95% credible interval is shown in Figure 31.

Next we consider the permeability field generated from the Gaussian random field as a reference spatial field which is taken as unknown in the model. The observed coarse-scale log permeability field, y_c, is calculated using the upscaling procedure in a 5×5 coarse grid. So the scale difference of the coarse scale permeability field with respect to the fine scale permeability field is 5 unit in each direction, i.e, each grid on coarse scale data is of 0.2×0.2 square unit. The fine-scale log permeability field is taken to be observed at 6 well locations along the boundaries, denoted by y_o. We use

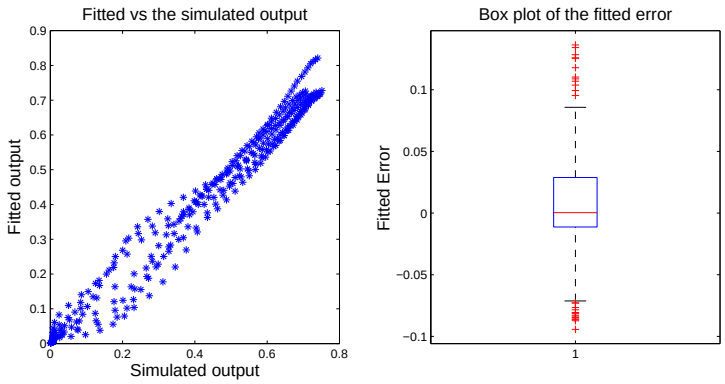

Fig. 30. Cross plot and box plot for the test data. Left: Cross plot of fitted vs simulated data, Right: Box plot of the residuals.

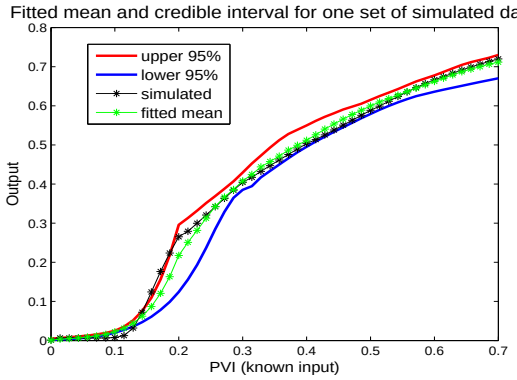

Fig. 31. Fitted mean and 95% credible interval for one set of test data using the emulator.

$n_r = 50$ observed real output data corresponding to the reference permeability field and $n_s = 5000$ simulation data corresponding to the 100 simulation runs in our model. We sample 200000 samples from the posterior by the hybrid reversible jump MCMC, Gibbs and Metropolis Hastings method described before. After 10000 burn in period we retain every 10th sample. Figure 32 shows the reference log permeability field and the mean of the posterior log permeabilty field. We can see that the posterior mean is very close to the reference log permeability field. Figure 33 shows the posterior density of the highest DCT coefficient, θ_1 and the posterior dendinty of σ_z^2. We can see that the posterior density of θ_1 has a peak near 26 which is close to the true highest DCT coefficeint(26.43) of the transformed reference field. The marginal one dimensional and two dimensional posterior distribution of the top four DCT coefficients are shown in Figure 34. The box plot of the marginal posterior of the DCT coefficients are shown in Figure 35. We also compare the computer efficiency in terms of CPU time for emulator based MCMC method to the regular simulator based MCMC method. The results are shown in Table III for the simulated example in a 25 grid and we can see the emultar based method is atleast 20 times faster than the simulator based method.

IV.4.2. Real Field Example

In this subsection we apply our model on a real field example , viz. punq-s3 model dataset. The PUNQ-S3 case has been taken from a reservoir engineering study on a real field example provided by Elf Exploration Production. It was qualified as a small-size industrial reservoir engineering model. The model contains $19 \times 28 \times 5$ grid blocks of which 1761 blocks are active. The PUNQ-S3 data set was an experimental study where the true permeability was actually known on the $19 \times 28 \times 5$ grid but the researchers were asked not to use the permeability data for their modeling purpose.

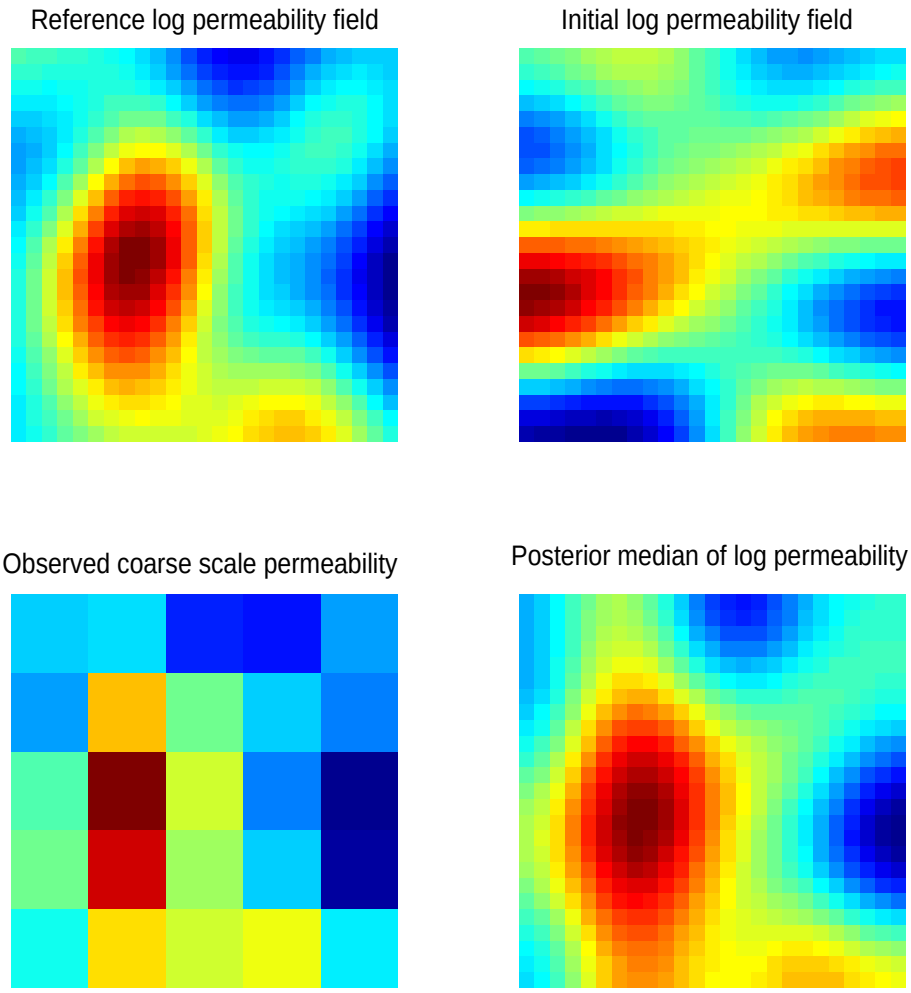

Fig. 32. Log permeability field for the simulated example using emulator. Top left: Reference log permeability field, Top right: Initial log permeability field of the Markov chain, Bottom Left: Observed coarse-scale permeability, Bottom right: Posterior median of the log permeability field.

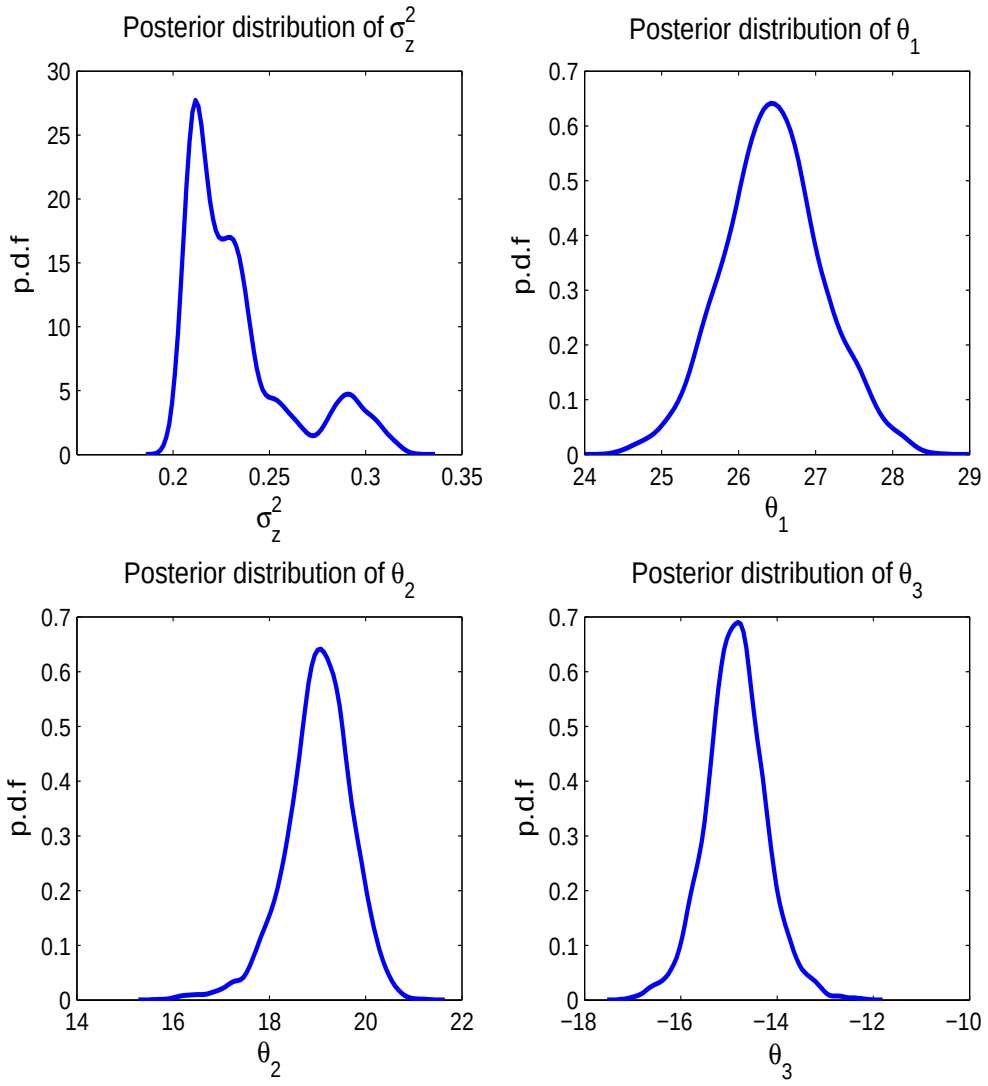

Fig. 33. Posterior distributions of the model parameters for the simulated example using emulator. Top Left: Posterior density of σ_z^2, Top Right: Posterior density of θ_1, Bottom Left: Posterior density of θ_2, Bottom Right: Posterior density of θ_3.

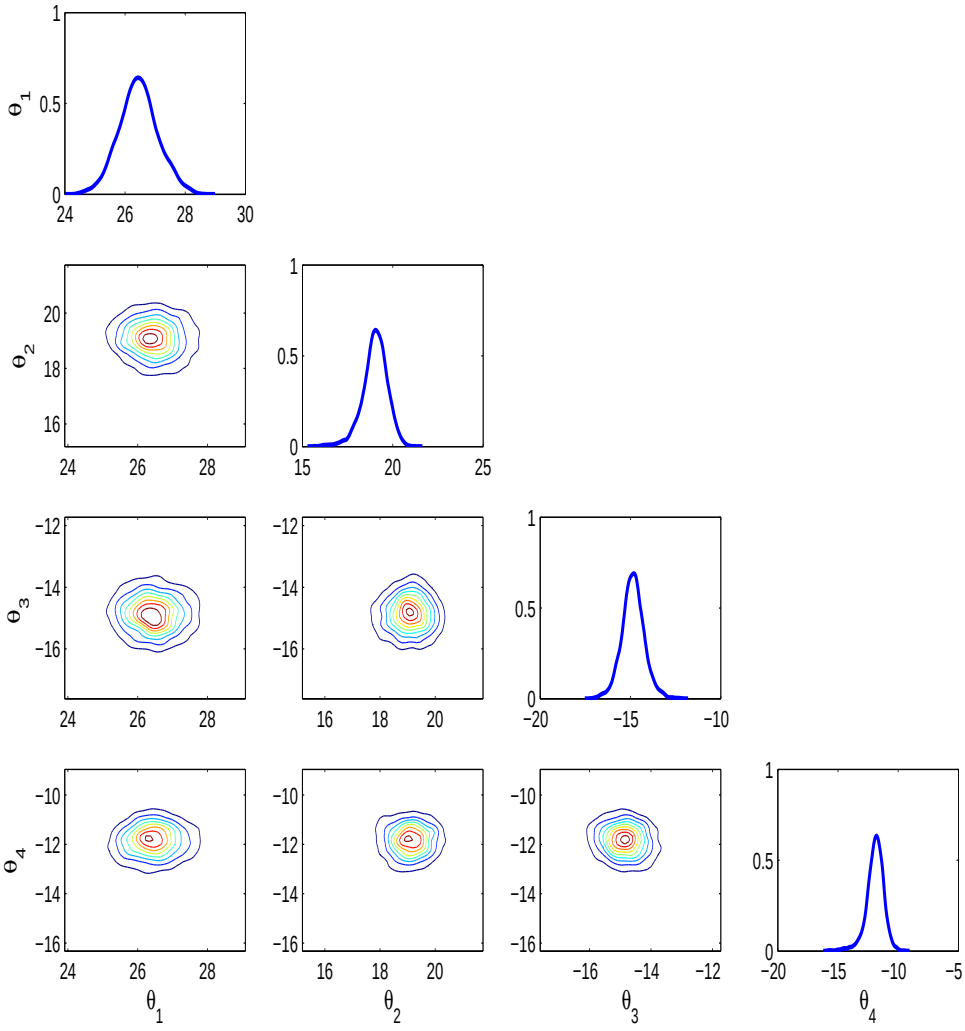

Fig. 34. One dimensional and two dimensional posterior marginals of the largest DCT coefficients for the simulated model.

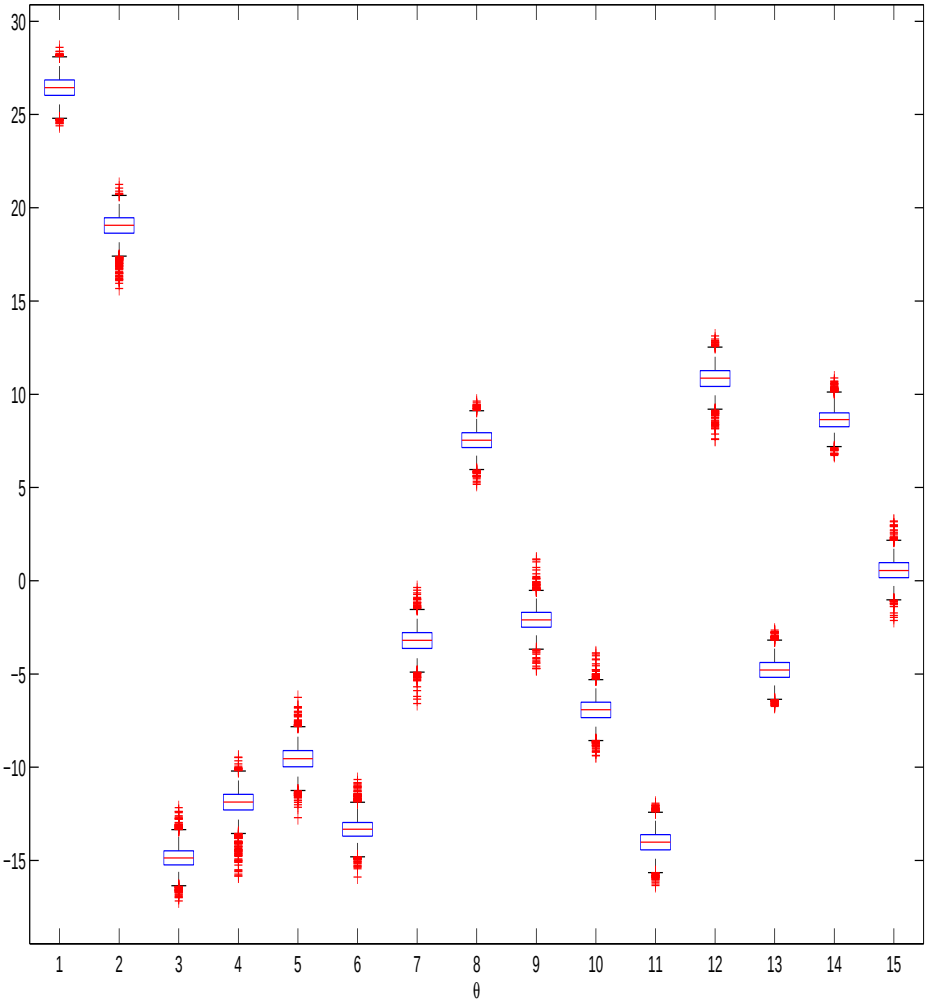

Fig. 35. Boxplot of the posterior marginals of the DCT coefficients for the simulated model.

Table III. Computational times, in seconds, of the emulator based and simulator based MCMC methods.

MCMC method	Time per likelihood calculation	Time per MCMC iteration	Total time for inversion (200000 samples)
Simulator based	5.000	5.112	1022400
Emulator based	0.112	0.212	42900

They were asked to use the production history only to infer about the true permeability field and then compare how their model resembles the true permeability field. For our example, we consider only the second most top layer of the five layers in the dataset and follow the same guidelines. We have used the 50 production history i.e the water-cut data, the permeability data on a 5×5 coarse grid and the true fine-scale permeability data only on the 6 well locations to infer about the fine-scale permeability field. The permeability measurements are expressed in the unit of mD where $1mD = 10^{-3}$ Darcy $= 10^{-12}m^2$. The spatial locations of the fine scale permeability field were given to the researchers in a transformed Cartesian co-ordinate system with each grid of 180×180 square unit starting from the origin, i.e co-ordinate of the top-left grid block is $(0, 0)$ and that of the bottom-right grid block is $(3420, 5040)$. Each grid on coarse scale data is of 684×1008 square unit. We use log transformation of the permeability data and logit transformation of the fractional flow data in our model. To built the BMARS emulator we again generate 100 samples of 16 DCT coefficients using Latin hypercube sampling with multivariate normal distribution. Each of these 16 DCT coefficients corresponds to a log permeability field obtained by the inverse

DCT transformation of a matrix whose elements of the top left square are those 16 DCT coefficients, the remaining elements of the matrix are set to zero. The other type of input considered are 50 injected pore volumes (rescaled to $0-1$) for each of these spatial fields. For each of these 5000 simulated input observations the output or water-cut data is simulated using computer codes corresponding to the forward simulator. We use the $n_r = 50$ observed real watercut data and $n_s = 5000$ simulated data in our model. We sample 200000 samples from the posterior distribution, after 10000 burn in period we retain every 10th sample.

From Figure 36 we can see that the posterior median of the sampled permeability field is close to the reference permeability field. The marginal posterior distribution of some of the model parameters are shown in Figure 37. Even though we have taken almost flat priors for the model parameters, from the posterior marginals we can see that how the observed data can reduce the uncertainties of the model parameters. The marginal one dimensional and two dimensional posterior distribution of the top four DCT coefficients are shown in Figure 38. The box plot of the marginal posterior of the DCT coefficients are shown in Figure 39. From the marginal distribution we can see that marginal posterior for all the DCT coefficients have a peak near the true value of DCT coefficients obtained by the DCT transformation of the reference log permeability field. Hence we can conclude that our Bayesian model can quantify the uncertainties in the unknown permeability field very well. The mean of the fitted output data corresponding to the reference permeability field together with its 95% credible interval are shown in Figure 40. From the credible interval and the mean plot we can conclude that the BMARS emulator can predict the output very well.

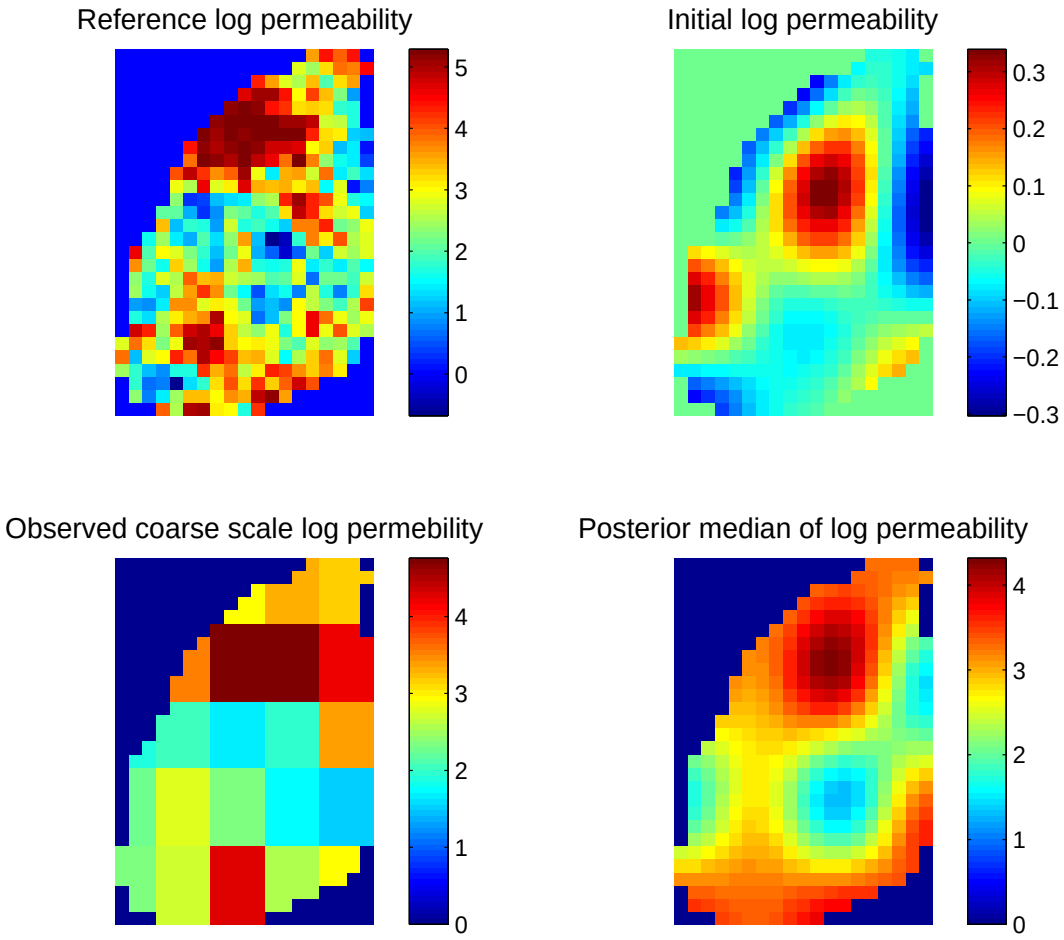

Fig. 36. Log permeability field for the PUNQ-S3 model using emulator. Top left: Reference log permeability field, Top right: Initial log permeability field of the Markov chain, Bottom Left: Observed coarse-scale permeability, Bottom right: Posterior median of the log permeability field.

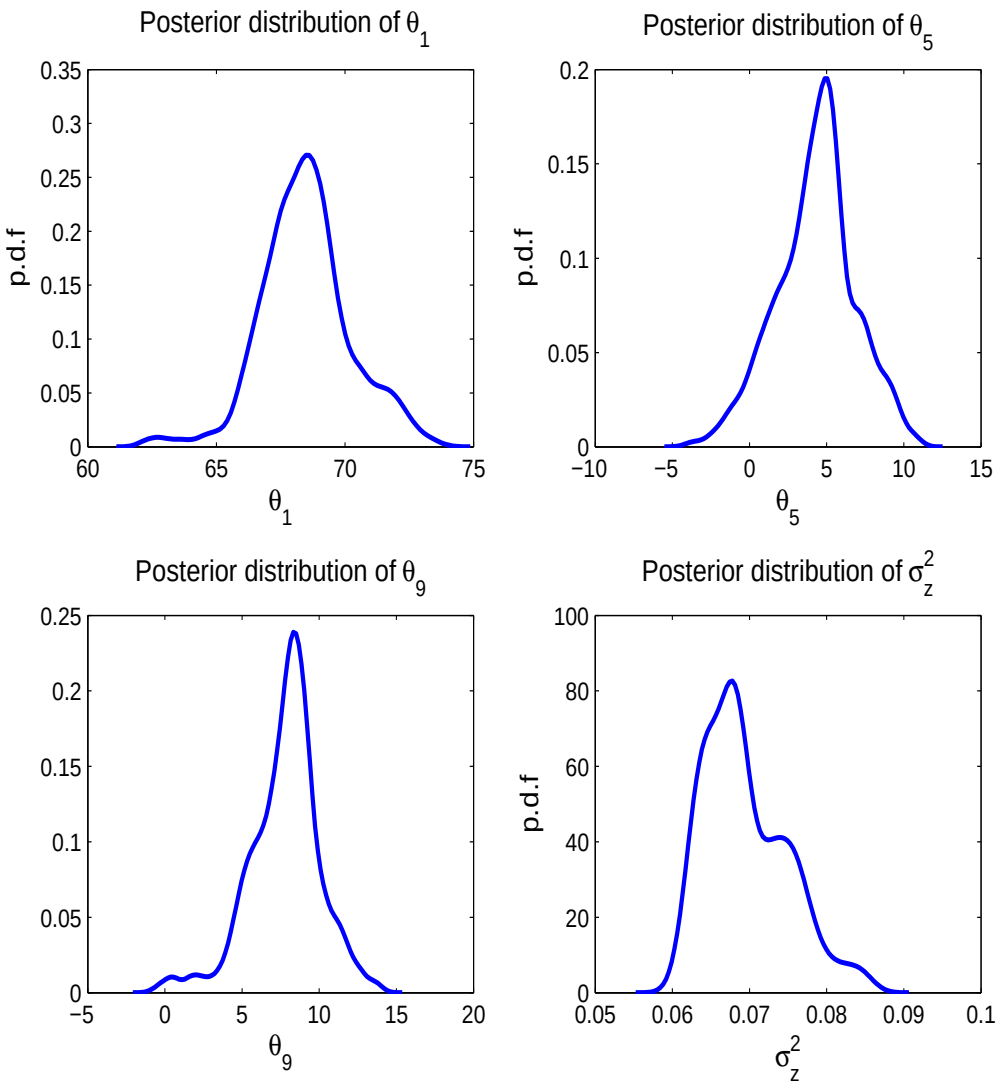

Fig. 37. Posterior distributions of the model parameters for the PUNQ-S3 model using emulator. Top Left: Posterior density of θ_1, Top Right: Posterior density of θ_5, Bottom Left: Posterior density of θ_9, Bottom Right: Posterior density of σ_z^2.

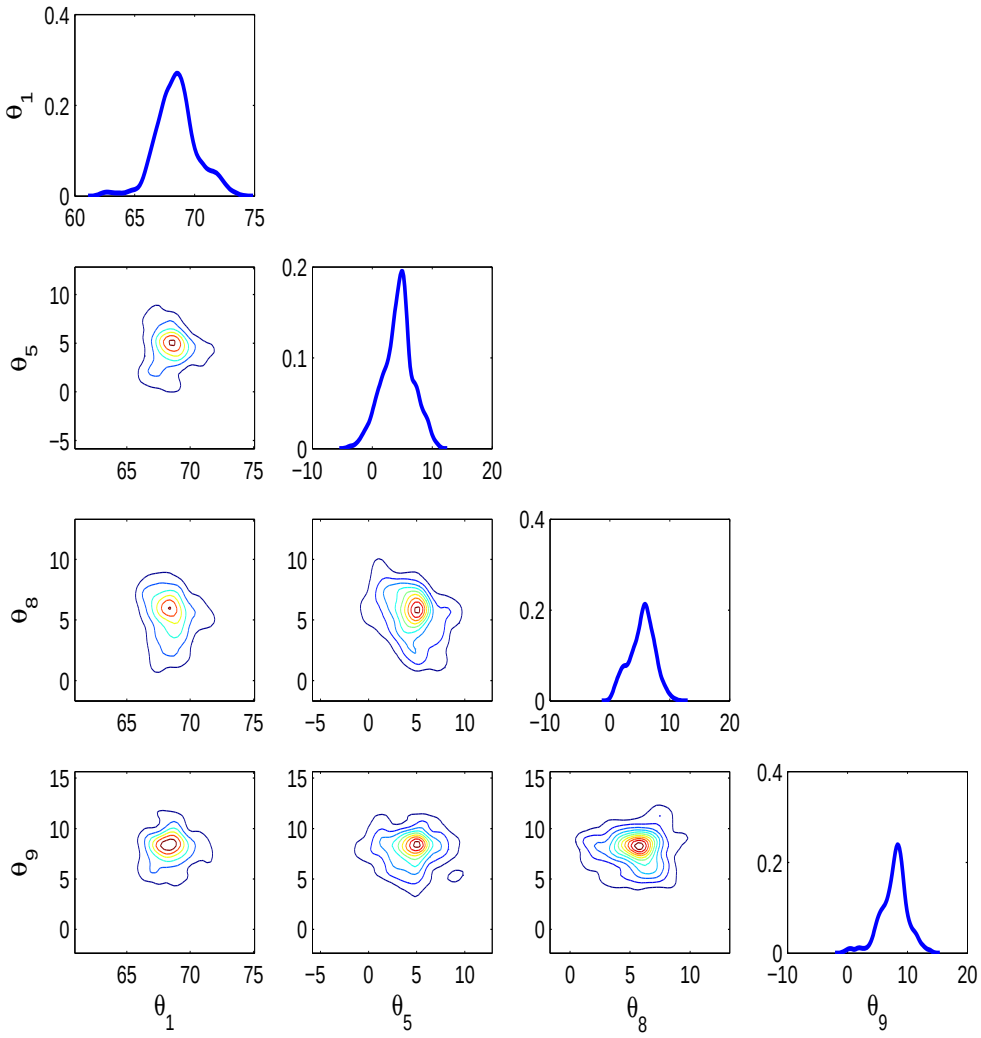

Fig. 38. One dimensional and two dimensional posterior marginals of the largest DCT coefficients for the PUNQ-S3 model.

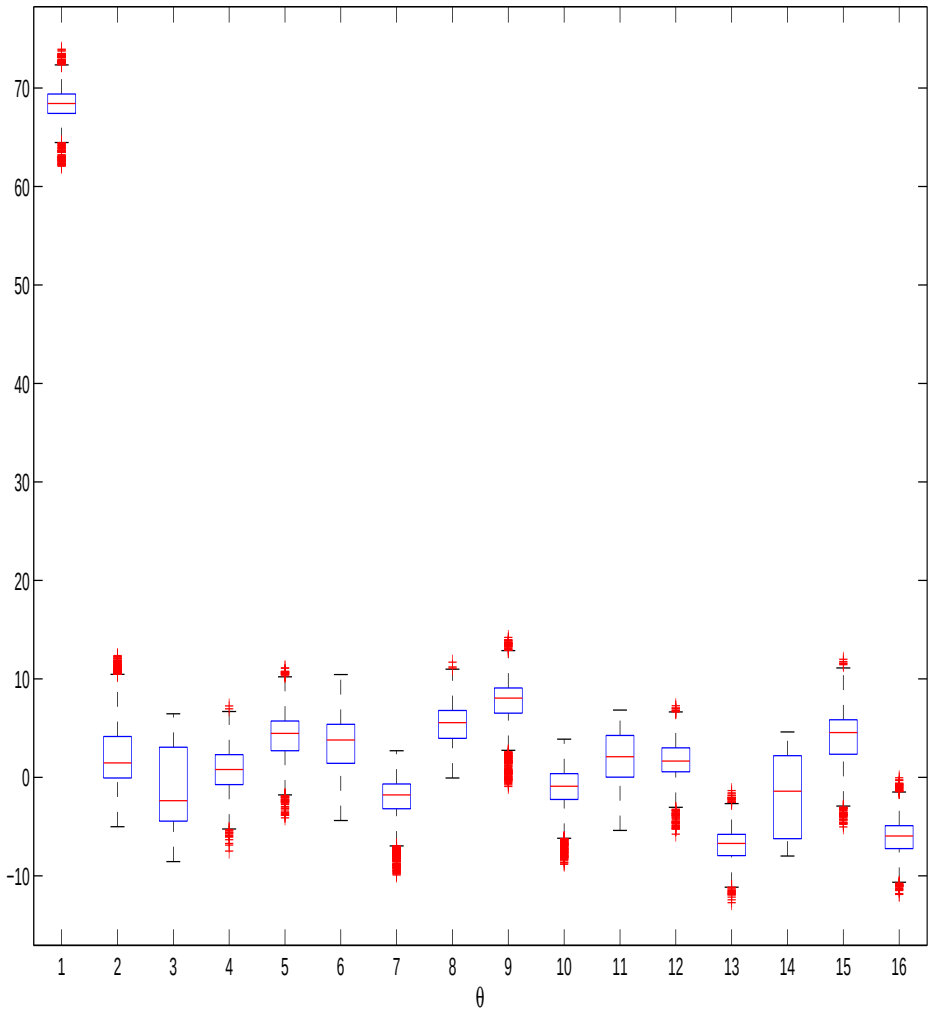

Fig. 39. Box plot of the posterior marginals of the DCT coefficients for the PUNQ-S3 model.

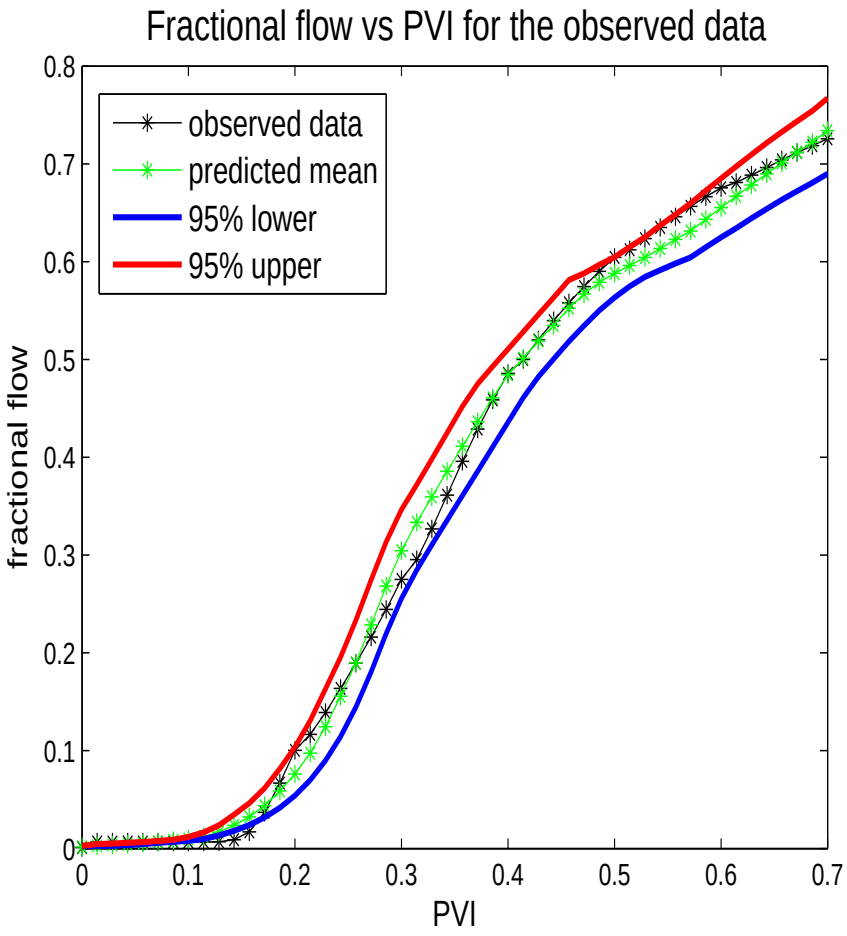

Fig. 40. Fitted mean and 95% credible interval of the observed output for the PUN-Q-S3 model using the emulator.

IV.5. Conclusion

The chapter pursues a Bayesian approach to inverse problems in which the unknown quantity is a spatial field. The posterior distribution provides a quantitative assessment of uncertainty in the inverse solution. The computational challenges associated with the repeated evaluation of the forward simulator is addressed. We use emulators based on Bayesian approach to multivariate additive regression splines to avoid the computational challenges of the direct simulation based approach. The unknown spatial field is parameterized by DCT transformation and the transformed DCT coefficients are used as regressors in the BMARS model. Numerical results shows that the BMARS emulator based MCMC method has substantial efficiency gain over the simulator based MCMC method in terms of CPU time. Our method is very flexible and can be applied to any physical process whose input is a spatial field. The method can be adapted to other inverse problems very easily as the mathematical model for the physical process was never used in the model. The only requirement is that the forward simulator on the spatial field has to be run on the design input points. Moreover the developed BMARS emulators can be easily used to prediction purpose which is very important in many fields such as production forecasting in oil reservoirs.

CHAPTER V

CONCLUSION AND SUMMARY

In this dissertation we have considered a Bayesian approach to solve nonlinear inverse problems where the unknown quantity is a random spatial field. We have used Bayesian hierarchical models to incorporate information from heterogeneous sources such as dynamic data and multiscale data. In order to reduce the computational challenges in solving the Bayesian inverse problems on a high dimensional spatial field we have used Karhunen-Loéve expansion and Discrete Cosine transform for dimension reduction of the random spatial field. In this Bayesian framework, we have shown that this inverse problem is well-posed by proving that the posterior measure is Lipschitz continuous with respect to the data in total variation norm. We have used Reversible Jump MCMC Algorithm where the number of leading terms retained in the K-L Expansion is also taken to be random and are updated by the data. We precompute the K-L expansion for a given set of the parameters and then use linear interpolation to find the respective eigen pairs for a given new value of the parameters at each MCMC step. Using Matrix perturbation theory we have shown that if the interpolating grid is small the approximated eigen pairs are very close to the true ones. The linear interpolation is very fast and thus saves us lot of CPU time. We have also developed two-stage reversible jump MCMC method which has the ability to screen the bad proposals in the first inexpensive stage and hence saves a lot of CPU time by rejecting the bad proposals very fast. Our numerical results from simulated and real field examples show that the proposed algorithm can accurately predict the unknown spatial field.

We have also studied the uncertainty quantification in inverse problems for heterogeneous spatial fields where the spatial fields have channelized structures. Chan-

nelized spatial fields are represented by facies boundaries and variogram-based spatial fields within each facies. Using level-set based approach, the shape of the channel boundaries is updated with dynamic data using a Bayesian hierarchical model where the number of points representing the channel boundaries is assumed to be unknown. Reversible jump Markov chain Monte Carlo algorithms are used to sample from the posterior distribution in such modeling. The search with Metropolis-Hastings algorithm results to very low acceptance rate, and consequently, the computations are CPU demanding. To speed-up the computations, we use coarse-scale models to screen the proposals. The truncation procedure of the K-L expansion introduce some error in the posterior measure. We also estimate a bound for the difference in the expectation of a square integrable function with respect to the full and the truncated posterior. Our computations show that the proposed algorithms are capable of capturing the channel boundaries and result to accurate predictions of subsurface properties.

We have developed statistical emulators based on Bayesian approach to multivariate additive regression splines (BMARS) over a large scale spatial field to avoid the expensive likelihood calculation, which contains the forward simulator, at each iteration of the MCMC step. To build the emulator at first the unknown spatial field is parameterized by Discrete Cosine Transformation (DCT) and then the transformed DCT coefficients are used as regressors in the BMARS model. The hybrid sampling technique, which is combination of reversible jump MCMC method, Metropolis Hastings method and Gibbs sampling method, is used to sample from the posterior. Numerical results shows that the BMARS emulator based MCMC method has substantial efficiency gain over the simulator based MCMC method in terms of CPU time.

For the definiteness of the problem, in this dissertation we have mainly considered the inverse problems from the reservoir characterization example but the developed

methodology to solve the Bayesian inverse problems are very flexible and can be efficiently applied to any physical process whose input is a high dimension spatial field.

REFERENCES

Ahmed, N., Natarajan, T. and Rao, K. R. (1974) Discrete cosine transform. *IEEE Trans. Comput.*, **C-23**, 90-93.

Barker, J. W. and Thibeau, S. (1997) A critical review of the use of pseudo-relative permeabilities for upscaling. *SPE Reservoir Engineering*, **12**, 138-143.

Bi, Z., Oliver, D. S. and Reynolds, A. C. (2000) Conditioning 3D stochastic channels to pressure data. *Soc. Petrol. Eng. J.*, **5**, 474-484.

Brigham E. O. (1988) *The Fast Fourier Transform and Its Applications*. Englewood Cliffs, NJ: Prentice Hall.

Caumon, G., Strebelle, S., Caers, J. and Journel, A. (2004), Assessment of global uncertainty for early appraisal of hydrocarbon fields. In *SPE 89943* presented at the *SPE Annual Technical Conference and Exhibition*, September, Houston, TX. Doi: 10.2118/89943-MS.

Christie, M. (1996) Upscaling for reservoir simulation. *J. Pet. Tech.*, **48**, 1004-1010.

Conti, S., Gosling, J. P., Oakley, J. E., O'Hagan, A, (2009) Gaussian process emulation of dynamic computer codes. *Biometrika*, **96**, 663-676

Cotter, S. L., Dashti, M., Robinson, J. C. and Stuart, A.M. (2009) Bayesian inverse problems for functions and applications to fluid mechanics. *Inverse Problems*, **25**, 115008, doi: 10.1088/0266-5611/25/11/115008

Cressie, N. (1993) *Statistics for Spatial Data*, rev. edn. New York: Wiley.

Datta-Gupta, A., King, M. J. (2007) *Streamline Simulation: Theory and Practise*. Richardson, TX: Society of Petroleum Engineers.

Denison, D. G. T., Mallick, B. K. and Smith, A. F. M. (1998) Bayesian MARS. *Statistics and Computing*, **8**, 337-346.

Dostert, P., Efendiev, Y., Hou, T. Y., Luo, W. (2006) Coarse-gradient Langevin algorithms for dynamic data integration and uncertainty quantification. *J. Comput. Phys.*, **217**, 123-142.

Dubrule, O. (1998) *Geostatistics in Petroleum Geology*, Continuing Education Course Notes, Number 38. Tulsa, OK: American Association of Petroleum Geology.

Durlofsky, L. J. (1998) Coarse scale models of two phase flow in heterogeneous reservoirs: Volume averaged equations and their relationship to the existing upscaling techniques. *Computational Geosciences*, **2**, 73-92.

Efendiev, Y., Datta-Gupta, A., Ginting, V., Ma, X. and Mallick, B. (2007) An efficient two-stage Markov chain Monte Carlo method for dynamic data integration. *Water Resour. Res.*, **41**, 1029-1035.

Efendiev, Y., Datta-Gupta, A., Osako, I., and Mallick, B. (2005) Multiscale data integration using coarse-scale models. *Advances in Water Resources*, **28**, 303-314.

Efendiev, Y., Hou, T. and Luo, W. (2006) Preconditioning Markov chain Monte Carlo simulations using coarse-scale models. *SIAM J. Sci. Comput.*, **28**, 776-803.

Efendiev, Y., Ma, X., Datta-Gupta, A. and Mallick, B. (2008) Multi-stage MCMC using non-parametric error estimators. Submitted to *Water Res. Resour.*

Egeland, T., Georgsen, F., Knarud R., Omre, H. (1993), Multifacies modelling of fluvial reservoirs. In *SPE 26502* presented at the *SPE Annual Technical Conference and Exhibition*, October, Houston, TX. Doi: 10.2118/26502-MS.

Ersoy, O. K. (1994) A comparative review of real and complex Fourier related transforms. *IEEE Proceedings*, **82**, 429-447.

Feig, E. and Winograd, S. (1992) Fast algorithms for the discrete cosine transform. *IEEE Transactions on Signal Processing*, **40**, 2174-2193.

Fienen, M. N., Kitanidis, P. K., Watson, D. and Jardine, P. (2004) An application of Bayesian inverse methods to vertical deconvolution of hydraulic conductivity in a heterogeneous aquifer at Oak Ridge National Laboratory. *Mathematical Geology*, **36**, 101-126.

Friedman, J. H. (1991) Multivariate adaptive regression splines. *Annals of Statistics*, **19**, 167.

Galli, A., Beucher, H., Le Loc'h, G. and Doligez, B. (1994) The pros and cons of the truncated Gaussian simulation, In *Geostatistical Simulations* (Eds.Armstrong, M., Dowd, P. A.), 217-233. Dordrecht, Netherlands: Kluwer Academic Publishers.

Gelman, A. and Rubin, D. B. (1992) Inference from iterative simulation using multiple sequences (with discussion and rejoinder). *Statistical Science*, **7**, 457-511.

Ghanem, R. G. and Spanos, P. D. (1991) *Stochastic Finite Elements: A Spectral Approach*. New York: Springer-Verlag.

Golub, G. H. and Van-Loan, C. F. (1996) *Matrix Computations*, 3rd edn. Baltimore, MD: Johns Hopkins University Press.

Gonzalez, R. C., Woods, R. E. (2002) *Digital Image Processing*, 2nd edn. Upper Saddle River, NJ: Prentice Hall.

Green, P. J. (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**, 771-732.

Haldorsen, H. and Damsleth, E. (1990) Stochastic modeling. *J. Pet. Tech.*, **42**, 404-412.

Higdon, D., Kennedy, M., Cavendish, J., Cafeo, J. and Ryne R. D. (2004) Combining field observations and simulations for calibration and prediction. *SIAM J. Sci. Comput.*, **26**, 448-466.

Jafarpour, B. and McLaughlin, D. B. (2007) Efficient permeability parameterization with the discrete cosine transform. In *SPE 106453* presented at the *SPE Reservoir Simulation Symposium*, February , Houston, TX. Doi: 10.2118/106453-MS.

Jain, A. K. (1989) *Fundamentals of Digital Image Processing*. Upper Saddle River, NJ: Prentice Hall.

Kaipio, J.P. and Somersalo, E. (2004) *Statistical and Computational Inverse Problems*. Applied Mathematical Sciences, New York: Springer-Verlag.

Kennedy, M. and O'Hagan, A. (2001) Bayesian calibration of computer models (with discussion). *J. R. Statist. Soc.* **B**, **63**, 425-464.

Kim, H. M., Mallick, B. K., Holmes, C. C. (2005) Analyzing nonstationary spatial data using piecewise Gaussian processes. *J. Am. Statist. Ass.*, **100**, 653-668.

Koltermann, C.E. and Gorelick, S. M. (1996) Heterogeneity in sedimentary deposits: A review of structure-imitating, process-imitating, and descriptive approaches. *Water Resour. Res.*, **32**, 2617-2658, 10.1029/96WR00025.

Landa, J.L. and Horne, R.N. (1997) A procedure to integrate well test data, reservoir performance history and 4-D seismic information into a reservoir description. In *SPE 38653* presented at the *SPE Annual Technical Conference and Exhibition*, October, San Antonio, TX. Doi: 10.2118/38653-MS.

Li, H. and Zhang, D. (2007) Probabilistic collocation method for flow in porous media: comparisons with other stochastic methods. *Water Resour. Res.*, **43**, W09409, doi:10.1029/2006WR005673.

Liu, N. and Oliver, D. S. (2005) Critical evaluation of the ensemble Kalman filter on history matching of geologic facies. *SPE Reservoir Evaluation and Engineering*, **8**, 470-477.

Loève, M. (1977) *Probability Theory*, 4th edn. Berlin: Springer.

Ma, X., Al-Harbi, M., Datta-Gupta, A. and Efendiev, Y. (2008) A Multistage sampling approach to quantifying uncertainty during history matching geological models. *Soc. Petrol. Eng. J.*, **13**, 77-87.

Marzouk, Y. and Najm, H. N. (2009) Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems. *J. Comput. Phys.*, **228**, 1862-1902.

Marzouk, Y., Najm, H. N. and Rahn, L.(2007) Stochastic spectral methods for efficient Bayesian solution of inverse problems. *J. Comput. Phys.*, **224**, 560586.

Marzouk, Y. and Xiu, D. (2009) A stochastic collocation approach to Bayesian inference in inverse problems. *Comm. Comput. Phys.*, **6**, 826-847.

McKay, M. D., Conover, W. J., and Beckman, R. J. (1979), A comparison of three

methods for selection values of input variables in the analysis of output from a computer code. *Technometrics*, **2**, 239-245.

Narasimha, M. J., Peterson, A. M. (1978) On the computation of discrete cosine transform. *IEEE Trans. Comm.*,**C-26**, 934-936.

Oakley, J. and O'Hagan, A. (2004) Probabilistic sensitivity analysis of complex models: a Bayesian approach. *J. R. Statist. Soc.* **B**, **66**, 751-769.

O'Hagan, A. (2006) Bayesian analysis of computer code outputs: a tutorial. *Reliability Engineering and System Safety*, **91**, 1290-1300.

Osher, S. and Fedkiw, R. (2003) *Level Set Methods and Dynamic Implicit Surfaces*. New York: Springer.

Pebesma, E. J. and Heuvelink, G. B. M. (1999) Latin hypercube sampling of Gaussian random fields. *Technometrics*, **41**, 303-312.

Ramsay, J. O. and Silverman, B. W. (2005) *Functional Data Analysis*, 2nd edn. New York: Springer.

Rao, K. R. and Yip, P. (1990) *Discrete Cosine Transform: Algorithms, Advantages, Applications*. Boston: Academic Press.

Reagan, M., Najm, H., Ghanem, R. and Knio, O. (2003) Uncertainty quantification in reacting flow simulations through non-intrusive spectral projection, *Combustion and Flame*. **132**, 545555.

Robert, C. and Casella, G. (1999) *Monte Carlo Statistical Methods*. New-York: Springer-Verlag.

Roberts, G. O. and Rosenthal, J. S. (2001) Optimal scaling for various Metropol-isHastings algorithms. *Statistical Science*, **16**, 351-367.

Ross, S. M. (1990) *A Course in Simulation*, New York: Macmillan.

Sacks, J., Welch, W., Mitchell, T. and Wynn, H. (1989) Design and analysis of computer experiments. *Statistical Science*, **4**, 409423.

Schwab, C. and Todor, R. (2006) Karhunen-Loève approcimation of random fields by generalized fast multipole methods. *J. Comput. Phys.*, **217** , 100-122.

Sethian, J. (1999) *Level Set Methods and Fast Marching Methods*. Cambridge, MA: Cambridge University Press.

Stein, M. (1987) Large sample properties of simulations using Latin hypercube sampling. *Technometrics*, **29**, 143-151.

Strinopoulos, T. (2005) Upscaling immiscible two-phase flows in an adaptive frame. *PhD Dissertation*. Applied and Computational Mathematics, California Institute of Technology, Pasadena, CA.

Tarantola, A. (2005) *Inverse Problem Theory and Methods for Model Parameter Estimation*. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Vogel, C. R. (2002) *Computational Methods for Inverse Problems*. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Waagepetersen, R. and Sorensen, D. (2001) A tutorial on reversible jump MCMC with a view toward applications on QTL mapping. *International Statistical Review*, **69**, 49-62.

Weber K.J. (1990) Influence of common sedimentary structures on fluid flow in reservoir models. *J. Pet. Tech.*, **34**, 665-672.

Williams, B., Higdon, D., Moore, L., McKay, M. and Keller-McNulty S. (2006) Combining experimental data and computer simulations, with an application to flyer plate experiments. *Bayesian Analysis*, **1**, 765-792.

Wong, E. (1971) *Stochastic Processes in Information and Dynamical Systems*. New York: McGraw-Hill.

Wu, X. H., Efendiev, Y., and Hou, T. (2002) Analysis of upscaling absolute permeability. *Discrete and Continuous Dynamical Systems, Series B*, **2**, 185–204.

Xiu, D. (2007) Efficient collocational approach for parametric uncertainty analysis. *Comm. Comput. Phys.*, **2**, 293-309

Xiu, D. and Hesthaven, J. (2005) High-order collocation methods for differential equations with random inputs. *SIAM J. Sci. Comput.*, **27**, 1118-1139.

Xiu, D. and Karniadakis, G. (2002) The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM J. Sci. Comput.*, **24**, 619-644.

Xiu, D. and Karniadakis, G. (2003) Modeling uncertainty in flow simulations via generalized polynomial chaos. *J. Comput. Phys.*, **187**, 137-167.

APPENDIX A

PROOF OF CONTINUITY OF THE POSTERIOR MEASURE

Proof of Theorem II.1.1

In this Appendix, we will show that the posterior measure is continuous with respect to the data in the total variation distance. For simplicity we consider the example of reservoir characterization (see Subsection I.1.1) with a layered permeability field, $k_f = k_f(x_2)$, and the flow along the layers. More precisely, we assume $p = 1$ at $x_1 = 1$ and $p = 0$ at $x_1 = 0$ and no flow on lateral boundaries. In this case, one can easily show that the velocity is given by $(k_f(x_2), 0)$. We consider the data $F_\tau(t) = \int_{out} v \cdot nSdl$. Then, $F_\tau(t) = \int_0^1 v(x_2)S(1, x_2, t)dx_2$. For simplicity, we assume that $S(x_1, x_2, t = 0) = S_0(x_1)$. In this case, $F_k(t) = \int_0^1 k_f(x_2)S_0(1 - k_f(x_2)t)dx_2$. To avoid sub-indices in the derivations, we replace x_2 by η, thus, $F_\tau(t) = \int_0^1 k_f(\eta)S_0(1 - k_f(\eta)t)d\eta$. Please note that in our notation $Y = log(k_f)$, $y_c = log(k_c)$, $y_o = log(k_f^o)$ and $y_p = log(k_f^p)$ where k_c is the observed coarse scale permeability field, k_f^o is the observed fine scale permeability at the well locations x^{obs} and k_f^p is the fine scale permeability field induced by K-L expansion on the well locations. The following notations will be used in the proofs and Lemmas.

$$\Psi_f(\tau, d) = \frac{||d - F_\tau||}{2} = \frac{\sum_{i=1}^{n} (d(t_i) - F_\tau(t_i))^2}{2}, \tag{A.1}$$

$$\Psi_f(\tau, y_c) = \frac{||y_c - L_\tau||}{2} = \frac{\sum_{i=1}^{N^*} (y_c(x_i) - L_\tau(x_i))^2}{2}, \tag{A.2}$$

$$\Psi_f(\tau, y_o) = \frac{||y_o - y_p||}{2} = \frac{\sum_{i=1}^{N_{obs}} (y_o(x_i^{obs}) - y_p(x_i^{obs}))^2}{2}. \tag{A.3}$$

Lemma 1. $\forall\ r > 0\ \exists\ C_1 = C_1(r)$ *such that if* $||d|| \leq r$, *then*

$$\Psi_f(\tau, d) \leq C_1\Big[\int_o^1 exp(2\sum_{l=1}^{m}\theta_l\Phi_l(\eta))d\eta + 1\Big]. \tag{A.4}$$

Proof.

$$\begin{aligned}
\Psi_f(\tau,d) &= 2^{-1}||d - F_\tau|| = \sum_{i=1}^{n}(d(t_i) - F_\tau(t_i))^2 \\
&\leq 2^{-1}\sum_{i=1}^{n}\Big[d(t_i) - \int_0^1 exp(\sum_{l=1}^{m}\theta_l\Phi_l(\eta))S_0(1 - exp(\sum_{l=1}^{m}\theta_l\Phi_l(\eta))t_i)d\eta\Big]^2 \\
&\leq 2^{-1}\sum_{i=1}^{n}\Big[2d(t_i)^2 + 2\Big(\int_0^1 exp(\sum_{l=1}^{m}\theta_l\Phi_l(\eta))d\eta\Big)^2\Big] \\
&\leq 2^{-1}\sum_{i=1}^{n}\Big[2d(t_i)^2 + 2\Big(\int_0^1 exp(2\sum_{l=1}^{m}\theta_l\Phi_l(\eta))d\eta\Big)\Big] \\
&= \Big[\sum_{i=1}^{n}d(t_i)^2 + n\Big(\int_0^1 exp(2\sum_{l=1}^{m}\theta_l\Phi_l(\eta))d\eta\Big)\Big] \\
&\leq C_1(r)\Big[1 + \Big(\int_0^1 exp(2\sum_{l=1}^{m}\theta_l\Phi_l(\eta))d\eta\Big)\Big],
\end{aligned} \tag{A.5}$$

where, $C_1(r) = \max(r^2, n)$. □

Lemma 2. $\forall\ r > 0\ \exists\ C_2 = C_2(r)$ *such that if* $||y_c|| \leq r$, *then*

$$\Psi_c(\tau, y_c) \leq C_2\left[\sum_i\Big(log\int_{\kappa_i} exp(\sum_{l=1}^{m}\theta_l\Phi_l(\eta))d\eta\Big)^2 + 1\right]. \tag{A.6}$$

Proof.

$$\begin{aligned}
\Psi_f(\tau,y_c) &= 2^{-1}||y_c - L_\tau|| = 2^{-1}\sum_{i=1}^{N^*}\Big[(y_c(x_i^c) - L_\tau(x_i^c))^2\Big] \\
&\leq 2^{-1}\Big[2\sum_{i=1}^{N^*}(y_c(x_i^c))^2 + 2\sum_{i=1}^{N^*}L_\tau(x_i^c)^2\Big] \\
&= \left[\sum_{i=1}^{N^*}(y_c(x_i^c))^2 + \sum_{i=1}^{N^*}\Big(log\int_{\kappa_i} exp(\sum_{l=1}^{m}\theta_l\Phi_l(\eta))d\eta\Big)^2\right]
\end{aligned}$$

$$\begin{aligned} &\leq \left[r^2 + \sum_{i=1}^{N^*}\Big(log \int_{\kappa_i} exp(\sum_{l=1}^{m} \theta_l \Phi_l(\eta))d\eta\Big)^2\right] \\ &= C_2(r)\left[1 + \sum_{i=1}^{N^*}\Big(log \int_{\kappa_i} exp(\sum_{l=1}^{m} \theta_l \Phi_l(\eta))d\eta\Big)^2\right], \end{aligned} \tag{A.7}$$

where, $C_2(r) = \max(r^2, 1)$. □

Lemma 3. *$\forall\ r > 0\ \exists\ C_3 = C_3(r)$ such that if $||y_o|| \leq r$, then*

$$\Psi_k(\tau, y_o) \leq C_3\Big[\sum_{l=1}^{m} \theta_l B_l + 1\Big]. \tag{A.8}$$

where, $B_l = \max_i (\Phi(x_i^{obs}))$

Proof.

$$\begin{aligned} \Psi_f(\tau, y_o) &= 2^{-1}||y_o - y_p|| = 2^{-1}\sum_{i=1}^{N_{obs}} (y_o(x_i^{obs}) - y_p(x_i^{obs}))^2 \\ &\leq 2^{-1}\sum_{i=1}^{N_{obs}} \Big[y_o(x_i^{obs}) - \sum_{l=1}^{m} \theta_l \Phi_l(x_i^{obs})\Big]^2 \\ &\leq 2^{-1}\sum_{i=1}^{N_{obs}} \Big[2y_o(x_i^{obs})^2 + 4\sum_{l=1}^{m} \theta_l \Phi_l(x_i^{obs})\Big] \\ &= \Big[\sum_{i=1}^{N_{obs}} y_o(x_i^{obs})^2 + \sum_{i=1}^{N_{obs}} 2\sum_{l=1}^{m} \theta_l \Phi_l(x_i^{obs})\Big] \\ &\leq \Big[r^2 + \sum_{i=1}^{N_{obs}} 2\sum_{l=1}^{m} \theta_l \Phi_l(x_i^{obs})\Big] \\ &\leq C_3(r)\Big[1 + \sum_{l=1}^{m} \theta_l B_l\Big], \end{aligned} \tag{A.9}$$

where, $B_l = \max_i(\Phi_l(x_i^{obs}))$ and $C_3(r) = \max(r^2, 2N_{obs})$. □

Lemma 4. *$\forall\ r > 0\ \exists\ C_4 = C_1(r)$ such that for every $d_1,\ d_2$ with $||d_1 \vee d_2|| \leq r$, we have*

$$|\Psi_f(\tau, d_1) - \Psi_f(\tau, d_2)| \leq C_4\Big[\int_o^1 exp(\sum_{l=1}^{m} \theta_l \Phi_l(\eta))d\eta + 1\Big]||d_1 - d_2||_2. \tag{A.10}$$

Proof.

$$
\begin{aligned}
|\Psi_f(\tau, d_1) - \Psi_f(\tau, d_2)| &= 2^{-1}\Big|\sum_{i=1}^{n}(d_1(t_i) - F_\tau(t_i))^2 - \sum_{i=1}^{n}(d_2(t_i) - F_\tau(t_i))^2\Big| \\
&\leq 2^{-1}\sum_{i=1}^{n}\Big|d_1(t_i) - d_2(t_i)\Big|\Big|d_1(t_i) + d_2(t_i) - 2F_\tau(t_i)\Big| \\
&\leq 2^{-1}\sum_{i=1}^{n}\Big|d_1(t_i) - d_2(t_i)\Big|\Big[\sum_{i=1}^{n}|d_1(t_i)| + \sum_{i=1}^{n}|d_2(t_i)| \\
&\quad +2\sum_{i=1}^{n}|F_\tau(t_i)|\Big] \\
&\leq ||d_1 - d_2||_{l_2}\Big[2r + \sqrt{2}\sum_{i=1}^{n}|F_\tau(t_i)|\Big] \\
&= ||d_1 - d_2||_{l_2}\Big[2r + \sqrt{2}\sum_{i=1}^{n}\Big|\int_0^1 exp(\sum_{l=1}^{m}\theta_l\Phi_l(\eta)) \\
&\times \quad S_0(1 - exp(\sum_{l=1}^{m}\theta_l\Phi_l(\eta))t_i)d\eta\Big|\Big] \\
&\leq ||d_1 - d_2||_{l_2}\Big[2r + \sqrt{2}n\Big|\int_0^1 exp(\sum_{l=1}^{m}\theta_l\Phi_l(\eta))d\eta\Big|\Big] \\
&\leq C_4\Big[\int_o^1 exp(\sum_{l=1}^{m}\theta_l\Phi_l(\eta))d\eta + 1\Big]||d_1 - d_2||_2, \qquad \text{(A.11)}
\end{aligned}
$$

where, $C_4(r) = \max(2r, \sqrt{2}n)$. □

Lemma 5. *$\forall\ r > 0\ \exists\ C_5 = C_5(r)$ such that for every $y_c^1,\ y_c^2$ with $||y_c^1 \vee y_c^2|| \leq r$, we have*

$$
|\Psi_c(\tau, y_c^1) - \Psi_c(\tau, y_c^2)| \leq C_2\Big[\sum_i\Big(log\int_{\kappa_i} exp(\sum_{l=1}^{m}\theta_l\Phi_l(\eta))d\eta\Big)^2 + 1\Big]||y_c^1 - y_c^2||_2. \quad \text{(A.12)}
$$

Proof. The proof is almost same as the technique used in Lemma 4 □

Lemma 6. *$\forall\ r > 0\ \exists\ C_6 = C_6(r)$ such that for every $y_o^1,\ y_o^2$ with $||y_o^1 \vee y_o^2|| \leq r$, we have*

$$
|\Psi_k(\tau, y_o^1) - \Psi_k(\tau, y_o^2)| \leq C_6\Big[\sum_{l=1}^{m}\theta_l B_l + 1\Big]||y_o^1 - y_o^2||_2. \qquad \text{(A.13)}
$$

Proof. The proof is almost same as the technique used in Lemma 4. □

Proof of Theorem II.1.1: First we show that Z is bounded above by a positive quantity.

$$\begin{aligned} Z = \int g(\tau, y)\pi_0(\tau)d\tau &\leq \frac{1}{b_f^{a_f+n/2}} \frac{1}{b_c^{a_c+N^*/2}} \frac{1}{b_k^{a_k+N_{obs}/2}} \int_\tau d\pi_0(\tau) \\ &= \frac{1}{b_f^{a_f+n/2}} \frac{1}{b_c^{a_c+N^*/2}} \frac{1}{b_k^{a_k+N_{obs}/2}}. \end{aligned} \tag{A.14}$$

Now we show that Z is bounded below by a positive quantity.

$$\begin{aligned} Z &= \int g(\tau, z)\pi_0(\tau)d\tau \\ &= \int \Big([b_f + \Psi_f(\tau, d)]^{(a_f+n/2)}[b_c + \Psi_c(\tau, y_c)]^{(a_c+N^*/2)}[b_k + \Psi_k(\tau, y_o)]^{(a_k+N_{obs}/2)}\Big)^{-1} d\pi_0(\tau) \\ &\geq \Big(\int [b_f + \Psi_f(\tau, d)]^{(a_f+n/2)}[b_c + \Psi_c(\tau, y_c)]^{(a_c+N^*/2)}[b_k + \Psi_k(\tau, y_o)]^{(a_k+N_{obs}/2)} d\pi_0(\tau)\Big)^{-1} \\ &\geq \Big[\int \Big(b_f + C_1[\int_o^1 exp(2\sum_{l=1}^{m} \theta_l \Phi_l(\eta))d\eta + 1]\Big)^{(a_f+n/2)} \\ &\times \Big(b_c + C_2[\sum_i \Big(log \int_{\kappa_i} exp(\sum_{l=1}^{m} \theta_l \Phi_l(\eta))d\eta\Big)^2 + 1]\Big)^{(a_c+N^*/2)} \\ &\times \Big(b_k + C_3[\sum_{l=1}^{m} \theta_l B_l + 1]\Big)^{(a_k+N_{obs}/2)} d\pi_0(\tau)\Big]^{-1}, \text{using Lemma 1, 2 and 3} \\ &\geq \Big[\Big(b_f + C_1[exp(2m\theta_M \Phi_M) + 1]\Big)^{(a_f+n/2)} \Big(b_c + C_2[\sum_i (\kappa_i + m\theta_M \Phi_M)^2 + 1]\Big)^{(a_c+N^*/2)} \\ &\times \Big(b_k + C_3[exp(m\theta_M B_M) + 1]\Big)^{(a_k+N_{obs}/2)} d\pi_0(\tau)\Big]^{-1}, \end{aligned} \tag{A.15}$$

where, $\theta_M = \max_l \theta_l$, $\Phi_M = \max_l \Phi_l$ and $B_M = \max_l B_l$.

So, we have,

$$\begin{aligned} |Z_1 - Z_2| &= \Big|\int g(\tau, z_1)d\pi_0(\tau) - \int g(\tau, z_2)d\pi_0(\tau)\Big| \\ &\leq \int \Big|g(\tau, z_1) - g(\tau, z_2)\Big| d\pi_0(\tau) \end{aligned}$$

$$= \int \Big| U_1V_1W_1 - U_1V_1W_1 \Big| d\pi_0(\tau), \text{ where } U_i = [b_f + \Psi_f(\tau, d_i)]^{(-a_f - n/2)},$$

$V_i = [b_c + \Psi_c(\tau, y_c^i)]^{(-a_c - N^*/2)}$ and $W_i = [b_k + \Psi_k(\tau, k_f^{io})]^{(a_k + N_{obs}/2)}, i = 1, 2.$

$$= \int \Big| (U_1 - U_2)V_1W_1 + (V_1 - V_2)U_2W_1 + (W_1 - W_2)U_2V_2 \Big| d\pi_0(\tau)$$

$$\leq \int \Big(\Big| (U_1 - U_2)V_1W_1 \Big| + \Big| (V_1 - V_2)U_2W_1 \Big| + \Big| (W_1 - W_2)U_2V_2 \Big| \Big) d\pi_0(\tau)$$

$$\leq \int \Big(\frac{\Big| [b_f + \Psi_f(\tau, d_1)]^{(a_f + n/2)} - [b_f + \Psi_f(\tau, d_2)]^{(a_f + n/2)} \Big|}{b_f^{2(a_f + n/2)} b_c^{(a_c + N^*/2)} b_k^{(a_k + N_{obs}/2)}}$$

$$+ \frac{\Big| [b_c + \Psi_c(\tau, y_c^1)]^{(a_c + N^*/2)} - [b_c + \Psi_c(\tau, y_c^2)]^{(a_c + N^*/2)} \Big|}{b_f^{(a_f + n/2)} b_c^{2(a_c + N^*/2)} b_k^{(a_k + N_{obs}/2)}}$$

$$+ \frac{\Big| [b_k + \Psi_k(\tau, y_o^1)]^{(a_c + N_{obs}/2)} - [b_k + \Psi_k(\tau, y_o^2)]^{(a_c + N_{obs}/2)} \Big|}{b_f^{(a_f + n/2)} b_c^{(a_c + N^*/2)} b_k^{2(a_k + N_{obs}/2)}} \Big) d\pi_0(\tau)$$

$$\leq C_1^* \int \Big(C_1 \Big[\int_o^1 exp(2 \sum_{l=1}^{m} \theta_l \Phi_l(\eta)) d\eta + 1 \Big] \Big| \Psi_c(\tau, d_1) - \Psi_c(\tau, d_2) \Big| \Big) d\pi_0(\tau)$$

$$+ C_2^* \int \Big(C_2 \Big[\sum_i \Big(log \int_{\kappa_i} exp(\sum_{l=1}^{m} \theta_l \Phi_l(\eta)) d\eta \Big)^2 + 1 \Big] \Big| \Psi_c(\tau, y_c^1) - \Psi_c(\tau, y_c^2) \Big| \Big) d\pi_0(\tau)$$

$$+ C_3^* \int \Big(C_3 \Big[\sum_{l=1}^{m} \theta_l B_l + 1 \Big] \Big| \Psi_c(\tau, y_o^1) - \Psi_c(\tau, y_o^2) \Big| \Big) d\pi_0(\tau)$$

$$\leq C_4^* ||d_1 - d_2||_2 + C_5^* ||y_c^1 - y_c^2||_2 + C_4^* ||y_o^1 - y_o^2||_2, \text{by Lemma 4, 5 and 6}$$

$$\leq B||z_1 - z_2||_2 \tag{A.16}$$

Thus,

$$\|\pi_1 - \pi_2\|_{TV} = \int \left| Z_1^{-1} g(\tau, z_1) - Z_2^{-1} g(\tau, z_2) \right| d\pi_0(\tau) \leq I_1 + I_2, \tag{A.17}$$

where,

$$I_1 = \frac{1}{Z_1} \int \Big| g(\tau, z_1) - g(\tau, z_2) \Big| d\pi_0(\tau), \tag{A.18}$$

$$I_2 = \frac{|Z_1 - Z_2|}{Z_1 Z_2} \int g(\tau, z_2) d\pi_0(\tau) \tag{A.19}$$

From (A.15) we obtain that Z is bounded below. From (A.16) and the fact that Z_1 is bounded below (see (A.15)), it follows that $I_1 \leq B_1(r)||z_1 - z_2||_2$. By the upper bound of Z in (A.14), and the lower bound of Z in (A.15) and also using the bound in (A.16) we have $I_2 \leq B_2(r)||z_1 - z_2||_2$. Thus, combining these results we have

$$\|\pi_1 - \pi_2\|_{TV} \leq C\|z_1 - z_2\|_2. \tag{A.20}$$

This completes the proof of Theorem II.1.1 for a simple layered permeability case in reservoir characterization. The proof can be extended to general inverse problem in spatial fields easily.

APPENDIX B

INTERPOLATION APPROXIMATION FOR KARHUNEN LOÉVE EXPANSION

Proof of Theorem II.2.1

Proof. For simplicity, let us denote A_l as A and $A_{l+\delta l}$ as $\tilde{A}$. Also let us denote $A = ((a_{i,j}))$ and $\tilde{A} = ((\tilde{a_{i,j}}))$. Suppose the eigen value decomposition of A and $\tilde{A}$ are given by:

$$A\Phi_i = \lambda_i \Phi_i \text{ and } \tilde{A}\tilde{\Phi}_i = \tilde{\lambda}_i \tilde{\Phi}_i, i = 1, 2...m. \tag{B.1}$$

First we show that $|\tilde{a_{i.j}} - a_{i,j}| \to 0$ as $|\delta l| \to 0$, $\forall (i,j)$.

$$\begin{aligned} |\tilde{a_{i,j}} - a_{i,j}| &= |exp(-\frac{d^2}{(l+\delta l)^2}) - exp(-\frac{d^2}{l^2})| \\ &= \Big|\Big[1 - \frac{d^2}{(l+\delta l)^2} + \frac{d^4}{2(l+\delta l)^4} - .....\Big] - \Big[1 - \frac{d^2}{l^2} + \frac{d^4}{2l^4} - \ldots\Big]\Big| \\ &= \Big|\Big(\frac{d^2}{l^2} - \frac{d^2}{(l+\delta l)^2}\Big) + \Big(\frac{d^4}{2(l+\delta l)^4} - \frac{d^4}{2l^4}\Big) + \ldots\Big| \\ &\to 0 \text{ as } |\delta l| \to 0. \end{aligned} \tag{B.2}$$

Now it follows from the Matrix Perturbation Theory (see Golub and Van-Loan (1996), Theorem 7.2.2) that $\tilde{\lambda}_i = \lambda_i + O(|\delta l|)$. Now suppose $\tilde{\Phi}_i = \Phi_i + \acute{\Phi}_i$, then we have

$$\begin{aligned} (A + \delta A)(\Phi_i + \acute{\Phi}_i) &= (\lambda_i + \delta\lambda_i)(\Phi_i + \acute{\Phi}_i) \\ \Rightarrow \Big(A + \delta A - (\lambda_i + \delta\lambda_i)\Big)\acute{\Phi}_i &= \delta\lambda_i\Phi_i + \delta A\Phi_i \\ \Rightarrow \acute{\Phi}_i &= O(|\delta l|). \end{aligned} \tag{B.3}$$

□

APPENDIX C

DETAILED BALANCED CONDITION OF TWO STAGE REVERSIBLE JUMP MCMC

Proof of Theorem II.4.1

In two stage reversible jump Markov Chain Monte Carlo method the final proposal ν is generated from the effective instrumental distribution

$$Q(\nu|\nu_n) = \alpha_p(\nu_n, \nu)q(\nu|\nu_n) + \Big(1 - \int \alpha_p(\nu_n, \nu)q(\nu|\nu_n)d\nu\Big)\delta_{\nu_n}(\nu). \tag{C.1}$$

The transition kernel of the Markov Chain ν_n generated by the two-stage reversible jump MCMC is given by,

$$K(\nu_n, \nu) = \alpha_f(\nu_n, \nu)Q(\nu|\nu_n), \text{for } \nu = \nu_n, \tag{C.2}$$

$$K(\nu_n, \nu_n) = 1 - \int_{\nu \neq \nu_n} \alpha_f(\nu_n, \nu)Q(\nu|\nu_n)d\nu. \tag{C.3}$$

So, the transition kernel is continuous when $\nu \neq \nu_n$ ans has positive probability for the event $\{\nu = \nu_n\}$. We have to show that the transition kernel satisfies the detailed balance condition

$$\pi(\nu_n)K(\nu_n, \nu) = \pi(\nu)K(\nu, \nu_n). \tag{C.4}$$

Proof. Equality in (C.4) is obviously true for $\nu = \nu_n$, For $\nu \neq \nu_n$,
$\pi(\nu_n)K(\nu_n, \nu) = \pi(\nu_n)\alpha_f(\nu_n, \nu)Q(\nu|\nu_n) = min\left(Q(\nu|\nu_n)\pi(\nu_n), Q(\nu_n|\nu)\pi(\nu)\right)$
$= min\left(\frac{Q(\nu|\nu_n)\pi(\nu_n)}{Q(\nu_n|\nu)\pi(\nu)}, 1\right) Q(\nu_n|\nu)\pi(\nu) = \alpha_f(\nu, \nu_n)Q(\nu_n|\nu)\pi(\nu) = \pi(\nu)K(\nu, \nu_n).$
Hence the proof follows. □

APPENDIX D

POSTERIOR ERROR INDUCED BY THE KARHUNEN LOÈVE TRUNCATION

Proof of Theorem III.3.1

Our goal is to estimate the difference in the expect value of a function with respect to two different posteriors, where one of them is a truncation of the other. We consider a simplistic setting as $D = [0,1] \times [0,1]$ with no interface in the permeability field, and assume that $\nabla p \in L_\infty(D)$, $k \in L_\infty(D)$ and $v \in L_\infty(D)$, where p is pressure, k is permeability field and v is the velocity associated with time of flight. First we notice that, the following lemmas hold, and estimations here are based on the the coupled PDE system (2.35)- (2.38).

Lemma 7. *$\|S_1 - S_2\|_{L_2} \leq C\|k_1 - k_2\|_{L_2(D)}$, where S_1 and S_2 are water saturations, provided that $S_i(x,y,t)$, $i = 1,2$ are Lipschitz with respect to all time.*

Proof. In order to get the estimation of saturations, we need the concept of time of flight. For a particle that starts at a point p at $t = 0$ and moves with velocity v, the flow map $P(p,T)$ is its position at time $t = T$,

$$\frac{dP}{dT} = v(P), P(p,0) = p.$$

Time of flight T characterizes particles motion under the velocity field, since velocity is a function of the spatial variable.

$$\frac{dT}{dP} = \frac{1}{v(P)},\ T = \int_p^P \frac{dr}{v(r)}.$$

Suppose S remains Lipschitz for all time, then by Strinopoulos (2005)

$$\begin{aligned}\|S_1 - S_2\|_{L_2(D)} &\leq C\|T_1 - T_2\|_{L_2(D)} \leq C\| \int_p^P \frac{dr}{v_1(r)} - \int_p^P \frac{dr}{v_2(r)}\|_{L_2(D)} \\ &\leq C\| \int_p^P \frac{v_2(r) - v_1(r)}{v_1(r)v_2(r)} dr\|_{L_2(D)} \leq C\|v_2 - v_1\|_{L_2(D)}, \qquad \text{(D.1)}\end{aligned}$$

since $v_1, v_2 \in L_\infty(D)$.

On the other hand, $v(x) = -k(x)\nabla p$, therefore,

$$\begin{aligned}\|v_1 - v_2\|_{L_2(D)} &= \|k_1\nabla p_1 - k_2\nabla p_2\|_{L_2(D)} \\ &\leq \|k_1\nabla(p_1 - p_2)\|_{L_2(D)} + \|k_1 - k_2\|_{L_2(D)}\|\nabla p_2\|_{L_\infty(D)} \\ &\leq \|k_1\nabla(p_1 - p_2)\|_{L_2(D)} + C\|k_1 - k_2\|_{L_2(D)}.\end{aligned}$$

Also we have $div(k_1\nabla p_1) = 0, div(k_2\nabla p_2) = 0$, then $div(k_1\nabla p_1) - div(k_2\nabla p_2) = 0$, and further $div(k_1\nabla(p_1 - p_2)) = div((k_2 - k_1)\nabla p_2)$, so

$$\begin{aligned}\|k_1\nabla(p_1 - p_2)\|_{L_2(D)} &= \|(k_2 - k_1)\nabla p_2\|_{L_2(D)} \leq \|k_1 - k_2\|_{L_2(D)}\|\nabla p_2\|_{L_2(D)} \\ &\leq C\|k_1 - k_2\|_{L_2(D)}, \text{ and so we have,}\end{aligned}$$

$$\|v_1 - v_2\|_{L_2(D)} \leq C\|k_1 - k_2\|_{L_2(D)}. \qquad \text{(D.2)}$$

Then from (D.1) and (D.2), we have $\|S_1 - S_2\|_{L_2(D)} \leq C\|k_1 - k_2\|_{L_2(D)}$. □

In Bayesian framework, the reference fractional flow or water-cut $F(k;t) = \int_0^t \int_0^1 v(1,y)S(1,\, y,t)dydt$ is matched to get the target posterior distribution.

Lemma 8. *$|F(k_1;t) - F(k_2;t)|^2 \leq C\|k_1 - k_2\|_{L_2(D)}$, where k_1 and k_2 are permeabilities, $F(k_1;t)$ and $F(k_2;t)$ are water-cut functions.*

Proof.

$$\begin{aligned} F(k;t) &= \int_0^t \int_0^1 v(1,y)S(1,y,t)dydt \\ &= \int_0^t [\int_0^1 v(1,y)S(1,y,t)dy - \int_0^1 v(0,y)S(0,y,t)dy]dt \\ &\quad + \int_0^t \int_0^1 v(0,y)S(0,y,t)dydt. \end{aligned}$$

We have $S(0,y,t) = 1$, $S_t + v \cdot \nabla S = 0$ then $\int_0^1 v(0,y)S(0,y,t)dy = \int_0^1 v(0,y)dy$ $= \int_0^1 v(s,y)dy$ for any $s \in [0,1]$, since v is divergence free.

$$\begin{aligned} F(k_1;t) &= \int_0^t [\int_{\partial D} v_1(x,y)S_1(x,y,t)dy]dt + \int_0^t \int_0^1 v_1(0,y)dydt \\ &= \int_0^t [\int_D div\{v_1(x,y)S_1(x,y,t)\}dxdy]dt + \int_0^t \int_0^1 v_1(s,y)dydt \\ &= \int_0^t [\int_D v_1(x,y) \cdot \nabla S_1(x,y,t)dxdy]dt + \int_0^t \int_0^1 v_1(s,y)dydt \\ &= \int_0^t [-\int_D (S_1)_t dxdy]dt + \int_0^t \int_0^1 v_1(s,y)dydt \\ &= -\int_D S_1(x,y,t)dxdy + \int_D S_1(x,y,0)dxdy + \int_0^t \int_0^1 v_1(s,y)dydt. \end{aligned}$$

There is similar result to $F(k_2;t)$, then

$$\begin{aligned} |F(k_1;t) - F(k_2;t)|^2 &= |\int_D (S_2(x,y,t) - S_1(x,y,t))dxdy \\ &\quad + \int_D (S_1(x,y,0) - S_2(x,y,0))dxdy \\ &\quad + \int_0^t \int_0^1 (v_1(s,y) - v_2(s,y))dydt|^2 \\ &\leq C(\int_D |(S_2(x,y,t) - S_1(x,y,t))|^2 dxdy \\ &\quad + \int_D |S_1(x,y,0) - S_2(x,y,0)|^2 dxdy \\ &\quad + \int_0^t \int_D |v_1(x,y) - v_2(x,y)|^2 dxdydt) \\ &\leq C\|k_1 - k_2\|^2_{L_2(D)}, \text{ by Lemma 7.} \end{aligned}$$

□

Now when the permeability field $k(x,\omega)$, described by a covariance matrix $R(x,y)$, has only one facies on D, we can write $\theta = (\theta_1, \cdots, \theta_N)$, then $\pi(\theta) \propto G(\theta_1, \cdots, \theta_N)\pi_0(\theta)$, $\widetilde{\pi}(\theta) \propto \widetilde{G}(\theta_1, \cdots, \theta_M)\pi_0(\theta)$, where, $\pi(\theta)$ is the posterior needed to be sampled, and $\widetilde{\pi}(\theta)$ is an approximation of $\pi(\theta)$ for computational simplicity. $\pi_0(\theta)$ is the prior distribution. $G(\theta_1, \cdots, \theta_N)$ and $\widetilde{G}(\theta_1, \cdots, \theta_N)$ are likelihoods, where

$$\begin{aligned} G(\theta_1, \cdots, \theta_N) &= \exp(-\frac{\int_0^T |F_{obs} - F(k_1(\theta_1, \cdots, \theta_N); t)|^2 dt}{\sigma_f^2}), \\ \widetilde{G}(\theta_1, \cdots, \theta_M) &= \exp(-\frac{\int_0^T |F_{obs} - F(k_2(\theta_1, \cdots, \theta_M); t)|^2 dt}{\sigma_f^2}). \end{aligned}$$

Lemma 9. $|G(\theta_1, \cdots, \theta_N) - \widetilde{G}(\theta_1, \cdots, \theta_M)| \leq \dfrac{C}{\sigma_f^2}\|k_1 - k_2\|_{L_2(D)}$.

Proof. Assume that $F(k_1;t)$ and $F(k_2;t)$, which are computed by forward model, are close to the observed data F_{obs}, then

$$\begin{aligned} & |G(\theta_1, \cdots, \theta_N) - \widetilde{G}(\theta_1, \cdots, \theta_M)| \\ \leq\ & \frac{C}{\sigma_f^2}|\int_0^T |F_{obs} - F(k_1;t)|^2 dt - \int_0^T |F_{obs} - F(k_2;t)|^2 dt| \\ \leq\ & \frac{C}{\sigma_f^2}(\int_0^T |2F_{obs} - F(k_2;t) - F(k_1;t)|^2 dt)^{\frac{1}{2}} \cdot (\int_0^T |F(k_1;t) - F(k_2;t)|^2 dt)^{\frac{1}{2}} \\ \leq\ & \frac{C}{\sigma_f^2}(\int_0^T |F(k_1;t) - F(k_2;t)|^2 dt)^{\frac{1}{2}} \leq \frac{C}{\sigma_f^2}\|k_1 - k_2\|_{L_2(D)}. \end{aligned}$$

□

Proof of Theorem III.3.1

Proof.

$$\begin{aligned}
|E_{\pi(\theta)}[f(\theta)] - E_{\tilde{\pi}(\theta)}[f(\theta)]| &\leq C\int |f(\theta)||G(\theta_1,\cdots,\theta_N) - \widetilde{G}(\theta_1,\cdots,\theta_M)|\pi_0(\theta)d\theta \\
&\leq \frac{C}{\sigma_f^2}\int |f(\theta)|\|k_1 - k_2\|_{L_2}\pi_0(\theta)d\theta \\
&\leq \frac{C}{\sigma_f^2}(\int |f(\theta)|^2\pi_0(\theta)d\theta)^{\frac{1}{2}}(\int \|k_1 - k_2\|_{L_2}^2\pi_0(\theta)d\theta)^{\frac{1}{2}} \\
&\leq \frac{C}{\sigma_f^2}(\int \|k_1 - k_2\|_{L_2}^2\pi_0(\theta)d\theta)^{\frac{1}{2}},
\end{aligned}$$

provided that $f(\theta)$ is square integrable with respect to Gaussian measure, e.g. polynomial functions. To estimate the error of truncation of K-L expansion, let $k_1 = \exp(\sum_{i=1}^{N}\theta_i\sqrt{\lambda_i}\psi_i)$, $k_2 = \exp(\sum_{i=1}^{M}\theta_i\sqrt{\lambda_i}\psi_i)$ and for simplicity we assume $\theta_i \overset{\text{iid}}{\sim} N(0,1)$, then

$$\begin{aligned}
&|\int f(\theta)\pi(\theta)d\theta - \int f(\theta)\widetilde{\pi}(\theta)d\theta|^2 \\
\leq\ & \frac{C}{\sigma_f^4}\int \|\exp(\sum_{i=1}^{N}\theta_i\sqrt{\lambda_i}\psi_i) - \exp(\sum_{i=1}^{M}\theta_i\sqrt{\lambda_i}\psi_i)\|_{L_2}^2\pi_0(\theta)d\theta \\
\leq\ & \frac{C}{\sigma_f^4}\int_D\int \exp(2\sum_{i=1}^{M}\theta_i\sqrt{\lambda_i}\psi_i)[1 - \exp(\sum_{i=M+1}^{N}\theta_i\sqrt{\lambda_i}\psi_i)]^2\pi_0(\theta)d\theta dxdy \\
\leq\ & \frac{C}{\sigma_f^4}\int_D I_1 I_2 dxdy,
\end{aligned}$$

where

$$\begin{aligned}
I_1 &= \int\cdots\int \exp(2\sum_{i=1}^{M}\theta_i\sqrt{\lambda_i}\psi_i)\pi_0(\theta_1,\cdots,\theta_M)d\theta_1\cdots d\theta_M \\
&= \prod_{i=1}^{M}\frac{1}{\sqrt{2\pi}}\int \exp(-\frac{1}{2}(\theta_i^2 - 2\sqrt{\lambda_i}\psi_i)^2 + 2\lambda_i\psi_i^2)d\theta_i = \exp(2\sum_{i=1}^{M}\lambda_i\psi_i^2)
\end{aligned}$$

,because ψ_i's are bounded and

$$\begin{aligned}
I_2 &= \int\cdots\int[1-\exp(\sum_{i=M+1}^{N}\theta_i\sqrt{\lambda_i}\psi_i)]^2\pi_0(\theta_{M+1},\cdots,\theta_N)d\theta_{M+1}\cdots d\theta_N \\
&= \int\cdots\int\{1-2\exp(\sum_{i=M+1}^{N}\theta_i\sqrt{\lambda_i}\psi_i)+\exp(2\sum_{i=M+1}^{N}\theta_i\sqrt{\lambda_i}\psi_i)\} \\
&\quad \prod_{i=M+1}^{N}\frac{1}{\sqrt{2\pi}}\exp(-\frac{\theta_i^2}{2})d\theta_i \\
&\leq 1-2(1+\frac{1}{2}\sum_{i=M+1}^{N}\lambda_i\psi_i^2)+1+2\sum_{i=M+1}^{N}\lambda_i\psi_i^2(\exp(2\sum_{i=M+1}^{N}\lambda_i\psi_i^2)+\frac{1}{2}) \\
&\leq C\exp(2\sum_{i=M+1}^{N}\lambda_i\psi_i^2)\sum_{i=M+1}^{N}\lambda\psi_i^2
\end{aligned}$$

Since k is a stationary spatial process on a bounded region, i.e a spatial process where the covariance function depends only on the distance not on the spatial location, so by, Schwab and Todor (2006) $\{\psi_i\}$ is uniform $L_\infty(D)$ bounded. So

$$\begin{aligned}
|\int f(\theta)\pi(\theta_1,\cdots,\theta_N)d\theta \quad - \quad \int f(\theta)\widetilde{\pi}(\theta_1,\cdots,\theta_N)d\theta| &\leq \frac{C}{\sigma_f^2}\{\int_D I_1I_2dxdy\}^{\frac{1}{2}} \\
&\leq C \quad \{\int_D\exp(2\sum_{i=1}^{N}\lambda_i\psi_i^2)\sum_{i=M+1}^{N}\lambda_i\psi_i^2dxdy\}^{\frac{1}{2}} \\
&\leq \quad C\{\sum_{i=M+1}^{N}\lambda_i\}^{\frac{1}{2}}.
\end{aligned}$$

□

VITA

ANIRBAN MONDAL

Department of Statistics

Texas A&M University

3143 TAMU

College Station, TX 77843-3143

c/o Bani K. Mallick, Ph.D.

EDUCATION

2011 Ph.D., Statistics, Texas A&M University

2007 Master of Science, Statistics, Michigan State University

2005 Master of Statistics, Statistics, Indian Statistical Institute, India

2003 Bachelor of Science, Statistics, University of Calcutta, India

RESEARCH INTERESTS

Bayesian hierarchical models, Uncertainty quantification, Spatial and spatio-temporal modeling, Bayesian geostatistics.